



Università degli Studi di Cagliari

**SCUOLA DI DOTTORATO IN**  
**Tossicologia-Oncologia e Patologia Molecolare**

XXIII Ciclo

**TITOLO TESI**

IDENTIFICAZIONE DEI FATTORI GENETICI COINVOLTI NELLA  
SUSCETTIBILITÀ ALLO SVILUPPO DEL TUMORE AL POLMONE

Settore disciplinare di afferenza: MED/04

Presentata da:

Dott.ssa Silvia Picciau

Direttore:

Prof. Gaetano Di Chiara

Relatore:

Prof.ssa Giovanna Maria Ledda

Esame finale anno accademico 2009 – 2010

## *Indice*

<b><i>Introduzione</i></b>	<b><i>1</i></b>
✓ <i>La genetica del cancro</i>	<i>2</i>
✓ <i>I polimorfismi a singolo nucleotide (SNPs)</i>	<i>11</i>
✓ <i>Le patologie complesse</i>	<i>15</i>
✓ <i>Gli studi di associazione genome wide</i>	<i>19</i>
✓ <i>Studi di associazione e DNA-Pooling</i>	<i>21</i>
✓ <i>Il tumore del polmone</i>	<i>22</i>
<b><i>Scopo della tesi</i></b>	<b><i>29</i></b>
<b><i>Materiali e Metodi</i></b>	<b><i>31</i></b>
• <i>Casistica</i>	<i>32</i>
✓ <i>Pazienti affetti da tumore polmonare</i>	<i>32</i>
✓ <i>Controlli sani di popolazione</i>	<i>32</i>
• <i>Metodiche</i>	<i>34</i>
✓ <i>Estrazione di DNA genomico da sangue mediante PAXgene Blood</i>	
<i>DNA kit (PreAnalytix, QIAGEN)</i>	<i>34</i>
✓ <i>Estrazione di DNA genomico da tessuto mediante QIAamp</i>	
<i>DNA Mini Kit (QIAGEN)</i>	<i>35</i>
✓ <i>Estrazione di RNA totale da tessuto mediante Rneasy Mini Kit (QIAGEN)</i>	<i>35</i>
✓ <i>Valutazione qualitativa e quantitativa degli acidi nucleici</i>	<i>36</i>
✓ <i>Creazione dei Pool di DNA</i>	<i>37</i>
✓ <i>Analisi Genome-wide sui Pool</i>	<i>38</i>
✓ <i>Scelta dei marcatori e disegno dei primer</i>	<i>40</i>

✓ <i>PCR (Polymerase-Chain-Reaction) e analisi di SNPs mediante Pyrosequencing</i>	40
✓ <i>RT-PCR (Reverse-Transcription-Polymerase-Chain-Reaction)</i>	42
✓ <i>Analisi di espressione dei geni TRIO e SNX9 mediante Real-Time PCR</i>	42
✓ <i>Analisi statistiche</i>	43
 <b><i>Risultati e Discussione</i></b>	 44
• <i>Analisi genome-wide mediante microarray</i>	45
• <i>Il locus 5p15.2</i>	53
✓ <i>Analisi del polimorfismo rs30616</i>	54
✓ <i>Il gene TRIO</i>	56
✓ <i>Analisi dell' espressione genica</i>	59
✓ <i>Analisi dell'espressione del gene TRIO</i>	60
• <i>Il locus 6q25.1-q26</i>	62
✓ <i>Analisi del polimorfismo rs724897</i>	63
✓ <i>Il gene SNX9</i>	65
✓ <i>Analisi dell'espressione del gene SNX9</i>	67
• <i>Ricerca delle isoforme dei geni TRIO e SNX9</i>	69
✓ <i>Le isoforme del gene TRIO</i>	69
✓ <i>Le isoforme del gene SNX9</i>	71
 <b><i>Conclusioni</i></b>	 72
 <b><i>Bibliografia</i></b>	 75

# ***Introduzione***

## ***La genetica del cancro***

Il cancro è una patologia caratterizzata da alterazioni dinamiche a carico del genoma (*Hanahan et al., 2000*). La tumorigenesi è un processo multifasico in cui la successione di alterazioni genetiche in grado di conferire un vantaggio di crescita guidano la progressiva trasformazione di una cellula normale in una cellula tumorale (*Renan MJ et al., 1993*). La crescita sregolata delle cellule tumorali deriva dall'acquisizione sequenziale di mutazioni somatiche e/o alterazioni dell'espressione di geni che ne controllano la crescita, la differenziazione e l'apoptosi o che garantiscono l'integrità del genoma, cioè in quei meccanismi che normalmente governano la proliferazione e l'omeostasi della cellula (*Foulds L. et al., 1954*)

Nel fenotipo delle cellule di tutti i tipi di tumore vengono riscontrate sei alterazioni essenziali:

1. autosufficienza nei segnali di crescita,
2. insensibilità ai segnali inibitori,
3. evasione dai normali meccanismi di morte cellulare programmata di apoptosi,
4. potenzialità di replicare in modo illimitato,
5. angiogenesi sostenuta,
6. potenziale capacità di invasione del tessuto circostante e metastasi (*Hanahan et al., 2000*).

Tutti i tumori hanno un'origine monoclonale, derivano cioè da un'unica cellula che ha subito nel proprio genoma una o più mutazioni. La progenie di tale cellula costituisce la popolazione cellulare neoplastica, che pur avendo la medesima origine, può andare incontro ad ulteriori mutazioni. L'acquisizione di nuove mutazioni può conferire un vantaggio replicativo, con la conseguente comparsa di subcloni con maggiori capacità aggressive (*Stratton et al., 2009*).

Il processo di cancerogenesi è guidato da mutazioni genetiche, epigenetiche e cambiamenti

nel trascrittoma. La persistenza di queste alterazioni può guidare l'attivazione di segnali e funzioni cellulari aberranti che producono cambiamenti premaligni, quali per esempio la displasia. Ulteriori alterazioni possono indurre angiogenesi e guidare l'invasione dei tessuti circostanti determinando la progressione dai primi stadi tumorali in cancro e infine in metastasi. Nelle fasi avanzate della malattia possono ancora essere individuati quelle alterazioni che sono avvenute negli stadi precoci. Nelle lesioni premaligne sono stati evidenziati cloni e subcloni cellulari con perdita dell'eterozigosità di geni necessari per l'omeostasi cellulare, instabilità dei microsatelliti e mutazioni a carico di particolari geni (come in p53 e nel recettore per i fattori di crescita dell'epidermide EGFR) (*Herbst et al., 2008*).

I danni a carico del genoma in grado di determinare l'acquisizione di un fenotipo neoplastico possono essere di tipo strutturale (mutazioni) o di tipo funzionale. I primi comportano la codificazione per prodotti proteici abnormi, che possono portare alla perdita della funzione fisiologica a cui sono normalmente deputati o ad esercitarla in maniera aumentata o ridotta. I secondi invece comportano una dis-regolazione in eccesso o in difetto dell'espressione genica. Le alterazioni nella normale sequenza del DNA possono essere di differenti tipi: sostituzioni in una singola base; inserzioni o delezioni di piccoli o grandi frammenti di DNA; riarrangiamenti (rottura del DNA e ricongiungimento ad un altro segmento in una differente posizione del genoma); amplificazione genica (aumento nel numero di copie talvolta anche di alcune centinaia di volte rispetto alle due normalmente presenti nel genoma diploide); riduzione del numero di copie, che può risultare nella completa assenza di una sequenza di DNA dal genoma della cellula tumorale (*Popescu NC et al., 1997*).

Inoltre la cellula tumorale può aver acquisito nuove sequenze di DNA da fonti esogene, come nel caso del papilloma virus, del virus Epstein Barr, o di quello dell'epatite B per i

quali è noto il contributo nella genesi di alcuni tipi di tumore (*Talbot SJ et al., 2004*).

Il genoma della cellula tumorale può acquisire anche alterazioni epigenetiche che possono incidere sull'espressione genica, come alterazioni nella struttura della cromatina. Tali alterazioni si manifestano con alterazioni nella metilazione del DNA e modificazione degli istoni. Le modifiche epigenetiche possono essere soggette agli stessi meccanismi di selezione naturale noti in genetica e vengono trasmesse stabilmente dalla cellula madre alla cellula figlia generando conseguentemente un differente fenotipo. Le aberrazioni epigenetiche, contrariamente alle mutazioni, sono reversibili e pertanto potenzialmente utilizzabili come target terapeutico (*Sharma S. et al., 2010*).

Infine non deve essere dimenticato che all'interno della cellula tumorale sono contenuti anche altri genomi. Infatti le migliaia di mitocondri presenti all'interno della cellula contengono ciascuno un genoma circolare di approssimativamente 17 kilobasi. Mutazioni somatiche nel genoma mitocondriale sono state riportate in molti tipi di tumore, tuttavia il loro ruolo nella cancerogenesi deve essere ancora chiarito (*Brandon M et al., 2006*).

Le mutazioni somatiche si verificano nel genoma di tutte le cellule in divisione, sia normali che neoplastiche. Esse possono essere il risultato di errori di incorporamento di un nucleotide durante la replicazione del DNA o dell'esposizione ad agenti mutageni endogeni o esogeni. Le mutazioni possono essere classificate da un punto di vista funzionale in base al rilievo che possono avere nello sviluppo del tumore. Vengono definite “*driver*” quelle mutazioni che conferiscono un vantaggio replicativo, che le porta ad essere selezionate positivamente durante l'evoluzione del tumore. Le altre mutazioni sono definite “*passengers*” poiché non conferiscono nessun vantaggio di crescita, e sono quindi considerate biologicamente neutre. È molto probabile che molti tumori portino più di una mutazione driver e che questo numero possa variare a seconda del tipo di tumore. La selezione biologica è esercitata principalmente sulle mutazioni non sinonime perché



queste sono in grado di alterare la struttura e la funzione della proteina. Al contrario le mutazioni sinonime sono biologicamente silenti e quindi non possono dar luogo ad un vantaggio selettivo (*Greenman C. et al., 2007*). Secondo il modello proposto da Greenman et al. il rapporto mutazioni non sinonime-mutazioni sinonime permetterebbe di distinguere le mutazioni driver da quelle passengers e quindi prevedere un maggior rischio di selezione di quelle cellule dotate di tali mutazioni e dotate di potenziale proliferativo tumorale. Un aumento della proporzione delle mutazioni non-senso implica maggior rischio di selezione svantaggiosa (*Haber DA and Settleman J., 2007*).

Viene generalmente ritenuto che le mutazioni driver si trovino in quei geni conosciuti come “geni del cancro” mentre le mutazioni passengers sarebbero distribuite casualmente all’interno del genoma. Tale considerazione deve però essere rivista poiché la maggior parte delle mutazioni presenti nelle cellule di un tumore sono di tipo passengers e si sono rivelate localizzate all’interno di quei geni operativi per la sopravvivenza del tumore. Appare quindi necessario stabilire dei tratti che siano in grado di discriminare tra le varie mutazioni di un tumore in modo da poter prevedere un approccio terapeutico rivolto alle alterazioni determinate da ogni specifica mutazione (*Bignell et al., 2010*).

Le mutazioni possono non solo determinare lo sviluppo di un tumore, ma incidere sulla sua progressione o conferire particolari caratteristiche alla popolazione cellulare neoplastica. Un’importante classe di mutazione driver è quella che conferisce resistenza alla chemioterapia. Alcuni tumori inizialmente responsivi alla terapia acquisiscono in un secondo momento caratteristiche che li rendono resistenti, facendo diventare inefficaci i trattamenti precedentemente utilizzati. Queste mutazioni potrebbero nascere come mutazioni passengers e nel corso della terapia trasformarsi in mutazioni driver, selezionando così un gruppo di cellule resistenti al trattamento chemioterapico, in grado di espandersi e determinare la ricomparsa della patologia (*Loeb LA et al., 2003*).

Sebbene la maggior parte delle mutazioni siano somatiche, circa l'1% dei tumori si sviluppa in individui appartenenti a famiglie in cui si verifica in modo ricorrente l'insorgenza di un particolare tipo di tumore. Tali individui presentano una mutazione non somatica, ma ereditata in linea germinale, in un particolare gene in ogni cellula del loro corpo. Queste mutazioni rendono le persone che ne sono affette particolarmente suscettibili all'insorgenza di una o più neoplasie poiché una mutazione somatica a carico del secondo e unico allele rimanente determina l'alterazione del gene colpito. Tuttavia è errato parlare di "tumori ereditari" ma è più opportuno riferirsi ad essi come "sindromi ereditarie di suscettibilità al cancro". Non è il tumore ad essere trasmesso secondo le classiche leggi mendeliane di trasmissione dei caratteri ereditari, ma sono le alterazioni genomiche, riguardanti uno o più geni, che vengono trasmesse e che costituiscono un elevato fattore di rischio per lo sviluppo di una o più neoplasie. Tali mutazioni rappresentano quindi la causa di iniziazione della cancerogenesi. Anche se rare, le sindromi ereditarie sono molto importanti dal punto di vista biologico. Attraverso lo studio di queste specifiche mutazioni si cerca di chiarire le origini molecolari dei tumori, sia di quelli sporadici che di quelli a carattere ereditario. La familiarità viene riconosciuta quindi come un'importante componente di rischio per lo sviluppo del cancro e l'identificazione dei geni specifici che contribuiscono a tale rischio è di fondamentale rilievo (*Fearon ER., 1997*).

Il termine "geni ereditabili del cancro" descrive quei geni che se colpiti da mutazioni incidono con un'alta penetranza nello sviluppo di sindromi tumorali. Tali mutazioni vengono trasmesse in genere in via germinale e la probabilità che un individuo che ne è colpito possa sviluppare il tumore è variabile e dipende da vari fattori: dal tipo di allele mutato, dall'età e dall'azione di altri geni chiamati modificatori. Il peso di altri fattori come la dieta, lo stile di vita e i fattori ambientali rimane invece ancora da chiarire. La variante allelica dei geni di suscettibilità potrebbe invece essere associata con un aumentato rischio

di sviluppare tumori di tipo sporadico o sindromi per i quali non è nota una componente genica ad alta penetranza (*Fearon ER., 1997*).

Un tumore si sviluppa in seguito all'accumulo di mutazioni all'interno di particolari categorie di geni che vengono definiti pertanto "geni del cancro", il cancro può quindi essere definito una patologia genetica (*Vogelstein B et al., 2004*). I geni responsabili della cancerogenesi appartengono a tre differenti classi: "gatekeeper", "caretaker" e "landscaper". I geni gatekeeper regolano l'ingresso della cellula in un percorso di oncogenesi. Tale classe è composta da oncogeni e oncosoppressori. I geni caretaker hanno invece la funzione di mantenerne l'integrità del genoma e sono quindi in grado di controllare la tumorigenesi per via indiretta. Mutazioni a carico di questi geni possono condurre verso l'instabilità genomica, con un conseguente incremento di mutazioni a carico di tutti i geni, compresi quelli gatekeeper. Alterazioni nei geni landscaper invece non colpiscono direttamente la crescita cellulare ma generano un ambiente circostante anomalo che contribuisce alla trasformazione neoplastica delle cellule (*Kinzler KW and Vogelstein B., 1997*). Tuttavia l'alterazione di uno solo di questi geni non è sufficiente per dar luogo ad un tumore. Sono necessarie mutazioni a carico di diversi geni per consentire che il tumore progredisca e invada il tessuto circostante. Infatti il rischio di sviluppare un tumore dipende non solo da un'iniziale mutazione che determina l'inizio del processo di tumorigenesi, ma anche dalla successive mutazioni che ne promuovono la progressione (*Michor et al., 2004*).

I geni conosciuti come proto-oncogeni codificano per proteine che regolano la crescita cellulare. Vengono definiti oncogeni in seguito all'acquisizione di una mutazione attivante che conferisce loro la capacità di proliferare in modo incontrollato. All'interno della normale omeostasi cellulare le proteine codificate dai proto-oncogeni sono fattori di crescita (TGF $\alpha$ , EGF e PDGF) e loro recettori (RET, famiglia erb-B); trasduttori del

segnale (RAS); fattori di trascrizione (i prodotti della famiglia MYC, MYB, JUN e FOS); componenti del ciclo cellulare (cicline e CDK) e altre proteine coinvolte nel controllo della vita della cellula (Bcl-2 e MDM2). Le oncoproteine codificate dagli oncogeni hanno di solito le stesse funzioni delle loro controparti normali, tuttavia essendo espresse in modo anormale, permettono alla cellula di evadere dai normali meccanismi di regolazione e crescere autonomamente (*Weinberg RA, 1996*).

Nel corso degli anni è stato evidenziato come gli oncogeni siano coinvolti in tutti gli stadi tumorali: nell'iniziazione, nella progressione, nell'angiogenesi e nella metastasi. Ciò può avvenire in seguito a due tipi di alterazioni: cambiamenti nella struttura del gene (l'oncoproteina prodotta ha funzione aberrante) o alterazioni nella regolazione dell'espressione genica (aumentata produzione della proteina che stimola la crescita cellulare). La continua espressione degli oncogeni è necessaria per mantenere in vita il tumore. Le mutazioni a carico degli oncogeni vengono pertanto definite attivanti poiché determinano un guadagno di funzione che conferisce un vantaggio selettivo alla cellula rispetto alle cellule circostanti. Tali mutazioni sono di carattere dominante: è sufficiente che ne venga colpito un solo allele perché il prodotto alterato possa esplicare eccessivamente la sua funzione (*Bishop JM, 1986*).

Tuttavia poiché il tumore possa proliferare in modo incontrollato è necessaria non solo la stimolazione operata dagli oncogeni ma anche la deprivazione degli opportuni sistemi di blocco che in condizioni normali limitano l'accrescimento cellulare.

Gli oncosoppressori sono geni codificanti per proteine che inibiscono la divisione cellulare. Le mutazioni a carico di questi geni determinano l'inattivazione o mancata sintesi di proteine e conseguente perdita dei limiti di proliferazione. Le mutazioni a loro carico (al contrario di quelle degli oncogeni) sono recessive, poiché è necessaria la mutazione inattivante di entrambi gli alleli per perdere la funzione esercitata. In una cellula normale

esistono due copie di ogni gene oncosoppressore. La mutazione che inattiva il primo allele può essere neutra, vantaggiosa o svantaggiosa. La prima mutazione è neutra se la copia rimanente dell'altro allele è sufficiente per svolgere la funzione di oncosoppressore, mentre viene definita svantaggiosa se l'oncosoppressore è coinvolto nel processo di apoptosi. In questo caso il prodotto proteico dell'oncosoppressore sfugge ai normali meccanismi di bilancio e determina l'attivazione del processo apoptotico. Infine, la prima mutazione viene definita vantaggiosa se l'oncosoppressore è aploinsufficiente e la presenza della copia rimanente dell'altro allele non è sufficiente a garantire le funzioni di oncosoppressore. La seconda mutazione invece è sempre vantaggiosa poiché la cellula ha entrambi gli alleli inattivati e ha quindi perso la funzione svolta dal gene. Alcuni oncosoppressori possono essere inattivati mediante ipermetilazione di sequenze promotore senza alcun cambiamento nella sequenza nucleotidica. La metilazione ha luogo a livello dei legami CpG e si verifica raramente nei tessuti normali (*Santarosa M. and Ashworth A., 2004*). Le proteine codificate dagli oncosoppressori sono fattori che inibiscono la crescita e i loro recettori (BRCA-1); molecole che regolano l'adesione cellulare (DCC e APC); regolatori della trasduzione del segnale (NF1 come disattivatore di RAS); regolatori del ciclo cellulare (Rb, p53, p21); regolatori dell'apoptosi (bcl-2 vs. bax e bad) (*Weiberg RA, 1996*).

Nel cancro l'attivazione degli oncogeni è spesso una conseguenza dell'amplificazione delle copie di un cromosoma, mentre l'inattivazione dei geni oncosoppressori è spesso causata da delezioni emizigoti associate a mutazioni o delezioni emizigoti. È quindi evidente che per la nascita di un tumore non sia sufficiente solo la perdita del controllo della crescita cellulare ma occorrono anche altri eventi per garantire la proliferazione, per organizzare il tessuto tumorale e consentire la formazione di una massa cellulare dotata di capacità invasiva e metastatica (*Weir et al., 2004*).

L'instabilità genetica è una caratteristica di molti tumori umani e pertanto è diventato uno dei campi più promettenti della ricerca in biologia molecolare. Sono stati identificati due differenti tipi di instabilità genomica (*Loeb LA., 1991*). In una piccola frazione di tumori del colon-retto e alcuni altri tipi di cancro, un difetto nel geni di sistema del riparo del DNA (Mismatch Repair) determina un incremento del tasso di mutazioni puntiformi e conseguentemente una diffusa instabilità dei microsatelliti (*Umar A et al, 1994*). I satelliti sono tratti cromosomici localizzati ai lati del centromero o nella parte terminale o anche interdispersi, caratterizzati dalla ripetizione successiva, per un certo numero di volte, di una sequenza di DNA definita "motivo". I satelliti sono deputati al mantenimento dell'integrità dei cromosomi ed al corretto appaiamento dei cromosomi omologhi nella meiosi e nella ricombinazione. Essi sono costituiti da sequenze ripetute in tandem con lunghezza variabile. Si possono distinguere i minisatelliti con lunghezza di circa 20000 kb e i microsatelliti che hanno invece una lunghezza di 400 bp. Quasi tutti i tumori del colon-retto e molti altri tipi di cancro presentano inoltre instabilità cromosomica (CIN), che si manifesta con un aumentato tasso di perdita o guadagno di interi cromosomi o di grandi parti di essi durante la divisione cellulare. L'instabilità cromosomica determina uno sbilanciamento nel numero di cromosomi (aneuploidia) con un conseguente aumento nel tasso di perdita dell'eterozigosità (LOH) e un incremento nel tasso di inattivazione degli oncosoppressori. I geni coinvolti nell'instabilità genomica svolgono la loro funzione biologica nella condensazione della cromatina, nella coesione dei cromatidi fratelli, nella struttura e funzione del cinetocore, nella formazione dei microtubuli e nei punti di controllo del ciclo cellulare. È stata inoltre ipotizzata l'esistenza di una connessione tra i geni mutati nelle sindromi ereditarie di aumentata suscettibilità al cancro, come quelli coinvolti nei meccanismi di riparo del DNA, e l'instabilità genomica. Deve essere altresì delucidato se l'instabilità genomica sia un evento precoce della tumorigenesi e quale sia il

suo peso nella progressione del tumore (*Lengauer C. et al., 1998*).

### ***I polimorfismi a singolo nucleotide (SNPs)***

I polimorfismi a singolo nucleotide (SNPs) sono variazioni in una singola base nella sequenza del DNA presenti in almeno l'1% della popolazione. Sono le varianti genetiche più comuni all'interno del genoma e sono quindi diventati i marcatori più frequentemente utilizzati nelle mappe degli studi di associazione. Si definisce quindi come polimorfismo genetico la presenza di alleli multipli per lo stesso locus. Anche le mutazioni dalle quali hanno origine le malattie ereditarie hanno creato uno o più alleli patologici per un gene normale. Si parla però di polimorfismo quando anche la varietà meno frequente del carattere polimorfo appare stabile nella popolazione, con una frequenza superiore a quella delle mutazioni ricorrenti in modo casuale: in pratica quando la classe corrispondente all'allele meno frequente sia presente nella popolazione considerata in una proporzione superiore all'1% (*Brookes AJ., 1999*). Secondo alcune stime, nella popolazione mondiale, si trovano circa dieci milioni di SNPs per i quali entrambi gli alleli sono presenti con una frequenza superiore all'1%. Tali SNPs (common SNPs) costituiscono il 90% della variabilità nella popolazione mondiale, mentre il restante 10% è costituito da un insieme diversificato di altre varianti rare (*Kruglyak L. and Nickerson DA, 2001*). Si stima che i polimorfismi siano distribuiti lungo il genoma ad intervalli non regolari, circa ogni 300bp. Si ipotizza che durante l'evoluzione interi gruppi di polimorfismi siano stati selezionati in regioni funzionali del genoma. Recentemente la posizione dei polimorfismi è diventata oggetto di grande attenzione. Polimorfismi non-sinonimi determinerebbero il cambiamento dell'aminoacido e quindi la sintesi di una proteina alterata. Polimorfismi nei promotori, nelle regioni di splice, nelle regioni UTR 3' e 5' e nelle regioni intrageniche potrebbero alterare l'espressione della proteina, attraverso modifiche nella sua regolazione, splicing

alternativo o stabilità dell'mRNA . Nonostante i polimorfismi intergenici siano comuni, solo pochi di essi hanno mostrato di poter alterare l'espressione genica (*Ameur A. et al., 2009*). È stato stimato che circa 50000-200000 polimorfismi avrebbero un effetto biologico (*Bernig T. and Chanock SJ.,2006*).

La formazione di una variante genetica si ha in seguito ad un singolo evento mutazionale che si manifesta con una frequenza molto bassa (nell'ordine di  $2.5 \times 10^{-8}$  mutazioni per sito o 175 mutazioni per genomi diploidi per generazione). Ogni nuova variante è inizialmente associata all'intera regione genomica che si trova nelle vicinanze del locus cromosomico in cui si è verificata la mutazione. Lo specifico insieme di alleli che si trovano sullo stesso cromosoma o nella medesima regione subcromosomiale è detto aplotipo, quindi l'insorgere di una mutazione porta alla formazione di un nuovo aplotipo, così come avviene la formazione di un aplotipo mosaico nel caso di ricombinazione fra due aplotipi diversi presenti sul cromosoma di origine materna e paterna (*Nachman MW, 2000*). Nel corso dell'evoluzione dell'umanità nei geni sono avvenute numerose mutazioni. Gli studi di associazione si basano sull'assunzione che una mutazione avvenuta in un singolo cromosoma ancestrale venga ereditata dalle generazioni future e trasmessa con l'intera regione genomica adiacente, come un unico blocco. Questa mutazione rappresenta quindi un segno distintivo, un marcatore, che permette di identificare quel particolare segmento del cromosoma ancestrale (*Pääbo S., 2003*). Risulta quindi che geni vicini vengono più facilmente coereditati in blocchi di aplotipi, risultando quindi associati tra popolazioni che originano da un antenato comune. Tale fenomeno viene definito linkage disequilibrium (LD) (*Reich DE and Lander ES, 2001*). Gli aploblocchi hanno quindi origine da corti segmenti di un cromosoma ancestrale passato attraverso numerose generazioni senza ricombinazioni. SNPs in forte linkage disequilibrium lungo il genoma vengono chiamati SNP-aploblocchi. In virtù di ciò non è necessario genotipizzare tutti gli



SNPs all'interno di un aploblocco, ma ne saranno sufficienti solo alcuni chiamati tagSNPs, attraverso i quali tutti gli altri SNPs potranno essere dedotti (*Cheng JC., 2007*). Si assume infatti che il genotipo dei polimorfismi non direttamente genotipizzati possa essere correlato con uno o più marcatori. Questa correlazione non casuale indica che tali marcatori sono in linkage disequilibrium tra loro (*Carlson et al., 2004*). Poiché la probabilità di ricombinazione fra due SNPs aumenta con l'aumentare della distanza fisica fra i due, si evince che il grado di associazione fra SNPs diminuisce progressivamente con la distanza (*Reich DE and Lander ES., 2001*). Tuttavia dopo la trasmissione attraverso varie generazioni i marcatori subiscono varie ricombinazioni genetiche e il linkage disequilibrium decade. Di conseguenza solo quei marcatori in stretta vicinanza di un particolare locus, per esempio il locus di un gene di predisposizione per una particolare patologia, rimarranno in linkage disequilibrium solo nelle generazioni degli individui affetti dalla medesima patologia (*Pääbo, 2003*). Diversi studi hanno però evidenziato come il quadro di LD lungo il genoma sia molto complesso, variabile e difficilmente prevedibile solo attraverso dei modelli teorici (*Gabriel SB et al., 2001; Abecasis et al., 2001; Dawson et al., 2002*). I polimorfismi sono quindi dei marcatori genetici che possono essere usati nelle mappe dei tratti complessi di una malattia ma richiedono analisi correlate per stabilire le cause o il significato funzionale della variante. La situazione viene complicata dal fatto che molti polimorfismi sembrano essere specifici per popolazione: circa il 15% dei polimorfismi potrebbe differire nelle diverse popolazioni (*Berning T. and Chanock S., 2006*). Allo scopo di fornire una risorsa pubblica che permetta di accelerare la ricerca genetica nel 2002 è stato istituito il progetto "International HapMap". Il progetto HapMap si occupa di analizzare le varianti genetiche disperse sull'intero genoma in differenti gruppi di popolazione appartenenti a diverse etnie, con la creazione di vere e proprie mappe. L'obiettivo iniziale era quello di genotipizzare almeno un polimorfismo ogni 5 kb lungo il

genoma in 270 individui provenienti a 4 diverse aree geografiche. Nel 2005 sono stati pubblicati i risultati della fase I del progetto: la genotipizzazione di approssimativamente 1.3 milioni di polimorfismi. Nella fase II del progetto HapMap sono stati genotipizzati più di 2.1 milioni di polimorfismi negli stessi individui analizzati nella fase I. È stato stimato che le mappe create contengano approssimativamente il 25-35% dei 9-10 milioni di polimorfismi presenti nel genoma umano (frequenza allelica minore (MAF) $\geq$ 0.05), con una densità di circa un polimorfismo ogni kilobase. L'incremento della densità degli SNPs nella fase II ha permesso un ulteriore sviluppo nelle analisi sull'intero genoma utilizzate negli studi di associazione. Attualmente la fase II fornisce la risorsa più completa e disponibile per la selezione di tag-SNPs sull'intero genoma (*International HapMap Consortium*). Uno dei risultati più interessanti scaturiti da questo progetto è la suddivisione del genoma umano in regioni cromosomiche, dette "blocchi di aplotipi" o aploblocchi, all'interno delle quali gli SNPs hanno un alto grado di associazione. Questo significa che, all'interno di un aploblocco la variabilità di una popolazione è rappresentata solo da pochi aplotipi (*Patil et al., 2001*). Gli aploblocchi sono separati da regioni con un basso livello di LD ed un'alta diversità, che corrispondono probabilmente a "hotspot" di ricombinazione. La conoscenza della struttura di LD e della struttura in blocchi del genoma ha dei risvolti pratici molto importanti (*Johnson et al., 2001*) in quanto, analizzando solo alcuni SNPs opportunamente scelti (tag SNPs) è possibile conoscere le altre varianti genetiche presenti nelle vicinanze in modo tale da rappresentare tutti gli aplotipi presenti nella regione cromosomica. Sulla base di studi empirici è stato dimostrato che la maggior parte delle informazioni relative alla variabilità umana rappresentata dai dieci milioni di SNPs presenti nel genoma può essere raccolta analizzando da 200.000 fino a 1.000.000 di tag-SNPs (*Carlson et al., 2003; Goldstein et al., 2003*). Il progetto HapMap ha avuto un ruolo fondamentale nello sviluppo dei metodi per la progettazione degli studi di associazione

mediante analisi sull'intero genoma. Inoltre la commercializzazione di piattaforme che permettono una genotipizzazione più economica ha portato ad una nuova fase nella ricerca in genetica. Vista la disponibilità pubblica dei dati dell'HapMap molti gruppi hanno potuto integrare i propri risultati con quelli già presenti nel database aggiungendo così nuove informazioni. I database dell' HapMap sono diventati quindi un'efficace strumento per effettuare studi di associazione genome-wide su larga scala. Infatti, la diminuzione dei costi e l'incremento della densità dei polimorfismi nei pannelli di genotipizzazione permette di analizzare l'intero genoma per lo studio di patologie complesse, sostituendo tale approccio a quello comunemente utilizzato dell'analisi di un gene candidato. La possibilità di associare le varianti genetiche presenti sull'intero genoma con i fenotipi molecolari sarebbe un potente mezzo per studiare lo sviluppo di molte patologie (*International HapMap Consortium*).

### ***Le patologie complesse***

Nel corso degli ultimi decenni numerosi studi hanno cercato di identificare i fattori di predisposizione al cancro. Sono stati così identificati dei geni che, se mutati in linea germinale, sono in grado di conferire una predisposizione ad alta penetranza nello sviluppo di particolari tumori, nel retinoblastoma (*Knudson AG., 1993*), nella poliposi adenomatosa familiare o APC (*Bülow S. et al., 2004*) e nel tumore della mammella (*Ford D. et al., 1995*). Tali mutazioni sono però rare e vengono trasmesse in linea germinale alla progenie degli individui affetti rendendoli suscettibili allo sviluppo di particolari patologie che vengono pertanto definite sindromi familiari. Lo sviluppo dei tumori sporadici viene associato da anni a fattori ambientali, dietetici e allo stile di vita (*Wei EK. et al., 2010*). Tuttavia alcuni studi epidemiologici hanno dimostrato come i parenti di primo grado di pazienti affetti da tumore abbiano un rischio di sviluppare tale patologia superiore rispetto alla popolazione

generale (*Johns LE and Houlston RS., 2001; Pharoah PD et al 1997; Sellers T.A et al., 1992; Fernandez E et al., 1994*). Le analisi di linkage, generalmente usate per studi di ereditarietà, hanno tuttavia fallito nell'identificare i loci genetici associati ad un aumentato rischio di sviluppo di alcune patologie. Questo tipo di analisi si sono rivelate particolarmente adatte nello studio di patologie in cui variazioni in un singolo gene sono necessarie e sufficienti per determinarne lo sviluppo, come la fibrosi cistica o la malattia di Huntington. Esistono però patologie il cui sviluppo è associato a molteplici varianti genetiche, ognuna delle quali però concorre singolarmente con un basso rischio relativo. La suscettibilità allo sviluppo di queste patologie, definite complesse, viene conferita da diversi tipi di interazione: l'interazione tra molteplici loci del genotipo (epistasi), l'interazioni tra il genotipo e l'ambiente ma anche in seguito ad altri fattori come l'inattivazione del cromosoma x o per esclusione allelica (*Risch NJ., 2000*). Il termine *tratto complesso* si riferisce ad un fenotipo che non si mostra recessivo o dominante con un'ereditarietà classicamente mendeliana e che non è attribuibile ad un singolo locus genetico. In generale la complessità si manifesta quando la semplice corrispondenza tra genotipo e fenotipo crolla e per il medesimo genotipo possono manifestarsi differenti fenotipi o differenti genotipi possono manifestarsi nel medesimo fenotipo a causa dell'effetto di diversi fattori, per interazioni con altri geni o per fattori ambientali. Per queste patologie pertanto si ammette una base determinata da un complesso di fattori aventi influenze strutturali diverse (*Lander ES and Schork NJ, 1994* ). Quale sia il peso di ogni fattore è oggetto di opinioni discordanti. Alcuni studi indicano l'ereditarietà come fattore di suscettibilità di forte rilievo (*Risch, 2001*), altri invece sostengono che i fattori ambientali siano i maggiori responsabili dell'insorgenza di patologie complesse come il cancro (*Lichtenstein P. et al., 2000*) Altri studi ipotizzano invece uno scenario in cui diversi fattori concorrono simultaneamente allo sviluppo del tumore (*Ponder BA., 1991*;

*Dragani TA.et al., 1996 ; Peto J and Houlston RS., 2001*). I dati ad oggi ottenuti suggeriscono che si possa escludere con una certa sicurezza uno scenario in cui per nessun tipo di tumore solo uno o pochi loci possano spiegare il rischio genetico, indipendentemente dal tipo della variante genetica (SNP o CNV) o dalla frequenza (sia essa comune, rara o molto rara). Il rischio di sviluppare un tumore sarebbe invece associato ad un'**ereditarietà poligenica** in cui sarebbero coinvolte centinaia o migliaia di varianti genetiche. Si afferma quindi un modello teorico governato da una complessa architettura e molteplicità di loci genetici in cui mutazioni nei singoli geni e varianti genetiche presenti in diversi loci creerebbero nel tessuto normale una situazione favorevole allo sviluppo del tumore, appunto una predisposizione, tale rischio sarebbe poi modulato in modo diverso in ogni individuo da fattori ambientali conducendo allo sviluppo del tumore. Quale sia il fattore di penetranza o di rischio di ogni variante è ancora sconosciuto (*Galvan et al., 2010*).

Per l'individuazione dei geni responsabili dello sviluppo di queste patologie si deve ricorrere all'analisi del genoma, in zone limitate di esso o nella sua totalità. Le tradizionali analisi di linkage cercano all'interno del genoma di individui colpiti dalla patologia in esame, regioni con un numero di alleli in comune superiore a quello atteso. Ciò indica la presenza all'interno della regione, dell'allele di predisposizione alla patologia (*Carlson et al., 2004*). Le analisi di linkage su famiglie hanno una grande capacità nell' identificare patologie in cui un singolo allele conferisce un elevato rischio di suscettibilità e sono state utilizzate per quelle patologie che si trasmettono con un'ereditarietà tipicamente mendeliana. Per quelle patologie in cui invece diverse varianti genetiche conferiscono ognuna un basso rischio relativo vengono preferiti gli studi di associazione, in cui avviene il confronto tra le varianti genetiche di individui affetti dalla patologia e individui sani di controllo (*Risch NJ., 2000*). In tal caso la ricerca delle varianti genetiche può avvenire in modo diretto,

all'interno di un gene-candidato ipoteticamente coinvolto nella patologia (*Tabor HK et al., 2002*) oppure in modo indiretto sull'intero genoma quando non è noto, neanche ipoteticamente, il coinvolgimento di alcun gene (*Risch N and Merikangas K., 1996*). Rispetto alle analisi di linkage su famiglie gli studi di associazione presentano il vantaggio della semplicità del reclutamento di un largo numero di individui affetti dalla patologia senza legami parentali, particolarmente utile nel caso di malattie che si sviluppano in tarda età. Tuttavia per individui non imparentati le regioni genomiche comuni saranno più piccole, poiché sono intercorse numerose generazioni e altrettante ricombinazioni e variazioni dall'antenato comune più recente. Per analizzare tali regioni in uno studio di associazione sarà quindi necessaria una maggior densità di marcatori distribuiti sull'intero genoma rispetto alle analisi di linkage tradizionali, nell'ordine di centinaia di migliaia di marcatori (*Carlson et al., 2004*).

### ***Gli studi di associazione genome-wide***

Lo studio di associazione genome-wide (sull'intero genoma) viene definito dal National Institute of Health (NIH) come uno “Studio sulle varianti genetiche comuni disperse sull'intero genoma con lo scopo di identificare le associazioni tra le varianti genetiche e particolari caratteristiche fenotipiche osservabili” (*NIH, Federal Regist. 2007*). Gli studi di associazione più diffusi sono quelli casi-controllo, in cui avviene il confronto tra le frequenze alleliche delle varianti genetiche disperse sull'intero genoma di pazienti affetti dalla patologia e individui sani di controllo (*Manolio et al., 2006*). Per ottenere ciò è necessario determinare una densità ottimale di marcatori genetici per individuare il maggior numero di varianti comuni, in un numero di individui sufficiente a garantire che anche varianti di modesto effetto possano essere rilevate (*Skol AD, et al. 2007*). Un tipico studio di associazione genome-wide è composto da 4 fasi:

1. la selezione di un grande numero di individui affetti dalla patologia in esame o i tratti di interesse e un gruppo di popolazione di controllo con cui effettuare un paragone adeguato;
2. estrazione del DNA, genotipizzazione e risultati che assicurino un'elevata qualità;
3. test statistici per l'associazione tra SNPs e la patologia;
4. replicazione dell'associazione identificata in una popolazione indipendente e esami sperimentali delle implicazioni in cui le varianti identificate sono coinvolte.; (*Pearson TA. and Manolio TA, 2008*).

Se vengono seguiti principi epidemiologici consolidati, gli studi caso-controllo possono produrre validi risultati che, specialmente per patologie rare, non possono essere ottenuti in altro modo (*Libioulle et al, 2007*). Per la riuscita dello studio deve essere determinata una proporzione ottimale tra i partecipanti allo studio e il numero di polimorfismi genotipizzati: una proporzione di polimorfismi inferiore al 5% nella prima fase dello studio limita fortemente l'identificazione di associazioni (*Hoover RN., 2007*). Inoltre l'approccio agli

studi di associazione può essere problematico perché il grande numero di test statistici che si devono eseguire presenta un potenziale di risultati falsi-positivi senza precedenti (*Hunter DJ and Kraft P., 2007*). Esistono infatti alcune variabili con un alto potenziale confondente: associazioni spurie tra le varianti genetiche e la patologia (*Risch NJ., 2000*), differenze etniche all'interno della popolazione selezionata (*Thomas DC and Witte JS., 2002*) e assenza di riproducibilità dei risultati ottenuti. Quest'ultima in particolare viene comunemente attribuita alla stratificazione della popolazione, alle differenze fenotipiche e agli errori di genotipizzazione (*Frayling TM and McCarthy ML., 2007*). La struttura della popolazione, cioè la stratificazione della popolazione dovuta alla presenza di diverse etnie o gruppi di differenti origini geografiche, è proprio una delle maggiori variabili confondenti citata. La presenza di sottogruppi di popolazione può far sì che le differenze di frequenza allelica ottenute in uno studio caso-controllo siano dei falsi positivi. Pertanto la selezione dei pazienti e dei controlli deve avvenire all'interno della stessa popolazione, così come deve essere rispettata l'omogeneità tra i due sessi (*Wacholder S et al., 2002*). Inoltre l'utilizzo di particolari gruppi di controllo, come donatori di sangue invece, può essere problematico nell'analizzare potenziali varianti genetiche derivanti dall'esposizione a fattori ambientali (*Pearson TA and Manolio TA., 2008*). Esistono però anche potenziali falsi negativi, in quanto un risultato di linkage disequilibrium negativo in una determinata regione genomica non esclude che un gene di tale regione non sia invece associato con le varianti genetiche analizzate (*Risch NJ., 2000*). Un'altra causa di falsi-negativi può essere determinata inoltre dall'assenza di varianti genetiche di rilievo nella piattaforma di genotipizzazione, o l'assenza di polimorfismi di particolare importanza nella popolazione oggetto dello studio. L'aumento della densità di polimorfismi e l'eterogeneità delle popolazioni rappresentate nelle piattaforme ridurrebbe certamente il problema (*Pearson TA and Manolio TA., 2008*). Recentemente sono state messe in commercio alcune piattaforme



che consentono la genotipizzazione di più di un milione di varianti genetiche disperse sull'intero genoma (*Gunderson KL et al., 2006*). Le piattaforme di genotipizzazione comprendono da 500000 a un milione di polimorfismi ed è stato stimato che coprano dal 67% all'89% dei polimorfismi nella popolazione Europea e Asiatica e dal 46% al 66% delle varianti della popolazione Africana (*International HapMap Consortium, 2007*).

Uno studio dell'NCI suggerisce di eseguire ulteriori studi per replicare i risultati ottenuti con un maggior numero di campioni anche se ciò spesso risulta problematico per la difficoltà di reperire un grande numero di individui affetti dalla patologia in un'unica popolazione (*Chanock SJ et al., 2007*). L'identificazione dei polimorfismi associati alla patologia attraverso studi di replicazione è il primo passo nell'identificare le varianti genetiche responsabili dello sviluppo della patologia in esame. Lo studio di associazione essenzialmente è in grado di identificare un locus genetico, ma fornisce poche informazioni sulle motivazioni del coinvolgimento del gene nella patogenesi. Queste informazioni dovranno essere cercate attraverso lo studio delle funzioni svolte dal gene, analisi di espressione genica o attraverso studi di manipolazione genetica in linee cellulari o modelli animali (*Pearson TA and Manolio TA., 2008*).

### ***Studi di associazione e DNA-Pooling***

Gli studi di associazione hanno l'inconveniente di avere dei costi molto elevati che li rende pertanto poco accessibili. Per ridurre i costi viene spesso utilizzato un approccio che viene definito DNA-pooling in cui identiche aliquote di DNA appartenente a diversi individui vengono unite a costituire un unico campione rappresentativo del gruppo in esame. L'analisi di varianti genetiche disperse sull'intero genoma effettuate sul pool può essere utilizzato come primo step nella realizzazione di uno studio di associazione (*Pearson JV. et al, 2007*). Il problema più comune che viene attribuito a questo approccio è il rischio di

perdere dati importanti o di incorrere in errori di misurazione che renderebbero i risultati ottenuti su pool diversi da quelli che si otterrebbero sui singoli campioni (*Barratt et al., 2002*). Tuttavia alcuni studi hanno dimostrato che attraverso l'uso di alcuni accorgimenti e l'applicazione di adatti algoritmi i pool di DNA possono essere utilizzati sia per lo studio di malattie monogeniche che di quelle complesse (*Pearson JV. et al., 2007; Zhang H et al., 2008; Macgregor et al., 2008; Sebastiani et al., 2008*). Infatti tale approccio è stato impiegato oltre che per la ricerca delle varianti genetiche di suscettibilità al cancro anche per lo studio di altre patologie complesse quali l'artrite reumatoide (*Steer et al., 2007*) e della schizofrenia (*Kirov et al., 2008*)

### ***Il tumore del polmone***

Il cancro del polmone è la prima causa di morte per patologie tumorali e si stima che sia responsabile ogni anno di più di un milione di morti (*Jemal et al., 2010*). Secondo i dati raccolti attraverso i registri dei tumori, il tumore del polmone rappresenta in Italia il 13.1% di tutti i tumori ma è quello con mortalità più alta (27.6%). In Sardegna invece la mancanza di un registro dei tumori rende impossibile una stima esatta dell'incidenza di questa patologia nell'isola (*AIRTUM Working Group, 2009*). Nonostante l'incidenza del tumore polmonare sia inferiore a quella del tumore della mammella nella donna e di quella del tumore alla prostata nell'uomo presenta una mortalità nettamente superiore, con sopravvivenza a 5 anni dalla diagnosi solo nel 15% dei casi. La diagnosi generalmente tardiva è sicuramente una delle cause responsabili dell'elevato tasso di mortalità (oltre il 75% dei tumori polmonari viene diagnosticato in pazienti che già presentano metastasi) inoltre esistono una grande varietà di sottotipi istologici del tumore che si manifestano con aspetti della malattia molto diversi tra loro. Contrariamente ad altri tipi di tumore, come

quelli alla mammella, alla prostata o al colon, per il tumore polmonare non esiste un programma di screening (*Sun S et al., 2007*).

Negli ultimi anni è stata osservata un'aumentata incidenza del tumore polmonare nelle donne rispetto agli uomini, e alcuni studi hanno suggerito che ciò fosse dovuto ad una maggiore suscettibilità alla patologia nel sesso femminile. Altri hanno invece ipotizzato che tale diversità fosse imputabile al costante aumento dell'abitudine al fumo tra le donne contrapposta ad una diminuzione tra gli uomini (*Dubey S, and Powell CA, 2007*).

Attualmente la terapia del tumore polmonare include la resezione chirurgica, la chemioterapia e la radioterapia (da sola o in associazione con la chemioterapia) a seconda dello stadio della patologia. Sfortunatamente, nonostante gli approcci terapeutici, la patologia è raramente curabile: presenta infatti una prognosi particolarmente sfavorevole e solo il 15% dei pazienti colpiti sopravvive più di 5 anni (*Jemal et al., 2006*).

Normalmente la diagnosi e la stadiazione del tumore avviene attraverso l'esame dell'escreato, la broncoscopia, l'agobiopsia percutanea del tessuto, la radiografia e inoltre ci si avvale anche dell'ausilio della tac 3D a spirale, considerata attualmente il mezzo più efficace. I sottotipi istologici vengono invece distinti attraverso l'analisi della morfologia cellulare: la diagnosi istopatologia è particolarmente importante poiché il trattamento differisce profondamente a seconda dell'isotipo del tumore. L'analisi del profilo di espressione genica del tumore sarebbe invece fondamentale per adottare strategie terapeutiche mirate verso i geni responsabili dello sviluppo della patologia. Ciò comporterebbe un miglioramento nell'esito della terapia stessa e nelle prospettive di sopravvivenza del paziente (*Bhattacharjee et al., 2001; Garber et al., 2001; Meyerson et al., 2004*).

Il tumore del polmone, sia esso correlato o no al fumo di sigaretta, possiede diversi profili molecolari, che includono le mutazioni e carico di p53, KRAS, EGFR e HER2 (*Herbst et*

*al.*, 2008). Sono state riscontrate numerose anomalie molecolari a carico sia di oncogeni che di geni oncosoppressori. L'amplificazione del DNA viene utilizzata come indicazione per individuare l'attivazione di un oncogene, mentre la perdita dell'eterozigosità (LOH) è spesso usata come indicatore di un'alterazione avvenuta a carico di un gene oncosoppressore (*Risch A and Plass C.*, 2008).

Le mutazioni a carico dell'oncogene *Kras* sono state le prime alterazioni molecolari individuate nell'adenocarcinoma. Le proteine *ras* appartengono ad una superfamiglia di proteine-G coinvolte nella regolazione della crescita cellulare. Sono state identificate mutazioni a carico di *Kras* nel 30% degli adenocarcinomi polmonari (*Rodenhuis S and Slebos RJ.*, 1992). La mutazione G→T nel codone 12 è stata riscontrata esclusivamente in pazienti con adenocarcinoma polmonare associato al fumo di sigaretta (*Ahrendt SA.et al.*, 2001). Alcuni studi hanno evidenziato che le mutazioni a carico di *Kras* siano predittive di resistenza chemioterapica e di una prognosi non favorevole (*Slebos RJ et al.*, 1990; *Sugio K. et al.*, 1992).

Tra le alterazioni molecolari presenti nell'adenocarcinoma grande interesse è rivolto verso le mutazioni dell'EGFR (recettore per il fattore di crescita dell'epidermide). Tali mutazioni sono associate con un esito sfavorevole della malattia ma vengono sfruttate per adottare una terapia mirata avvalendosi dell'utilizzo dei farmaci inibitori dell'attività tirosin-chinasica del recettore (erlotinib e gefitinib) (*Stahel R.A.*, 2007). L'EGFR appartiene alla famiglia dei recettori ErbB, della quale fanno parte anche HER2, HER3 e HER4. Sono stati identificati 11 ligandi di questo recettore, tra i quali il fattore di crescita dell'epidermide (EGF) e il fattore di trasformazione di crescita alfa (TGFα) (*Harris RC et al.*, 2003). Le mutazioni dell'EGFR vengono classificate in base alla sensibilità mostrata verso gli inibitori dell'attività tirosin-chinasica. La mutazione più frequente è la delezione "in-frame" nell'esone 19, seguono poi la mutazione puntiforme L858R nell'esone 21,

quella G719A/C nell'esone 18, associate con la risposta agli inibitori dell'attività tirosin-chinasica dell'EGFR. Inoltre sono state individuate tre mutazioni nel dominio tirosin-chinasico associate con la resistenza a tali farmaci, la più comune è quella nell'esone 20, T790M. Molti studi hanno mostrato un'associazione tra le mutazioni dell'EGFR e l'adenocarcinoma polmonare, l'assenza di una storia di abitudine al fumo, l'etnia asiatica e il genere femminile (*Shigematsu H. and Gazdar AF, 2006; Riely GJ et al., 2006*). In nessun tumore con EGFR mutato sono state riscontrate mutazioni di Kras, individuate invece nell'8% dei tumori. Ciò suggerisce che esistano due differenti vie molecolari che conducono al tumore polmonare, l'esposizione al fumo che risulta nei tumori con Kras mutato e l'esposizione ad un altro cancerogeno sconosciuto che conduce invece al tumore con l'EGFR mutato (*Kosaka T. et al., 2004*). Inoltre sono state riscontrate in adenocarcinomi di non fumatori overespressione del fattore di crescita HER2 e mutazioni attivanti a carico di RAF. Il gene RAF è mutato nel 2% degli adenocarcinomi e codifica per una chinasi treonina-serina specifica attivata dalla proteina Ras. Così come per l'EGFR anche queste mutazioni sono mutualmente esclusive con Kras (*Naoki K et al., 2002*).

Circa l'80-90% dei tumori polmonari sono attribuibili al fumo di sigaretta e approssimativamente solo il 10% dei tumori polmonari si sviluppa in soggetti non-fumatori. Il fumo di sigaretta rappresenta sicuramente il principale fattore di rischio ma anche altre sostanze sono fortemente associate con lo sviluppo di tumori polmonari. Studi epidemiologici indicano che l'esposizione all'asbesto, all'arsenico e al radon sia correlabile con lo sviluppo di questa patologia. Il coinvolgimento degli inquinanti atmosferici invece è fonte di discussione. Infatti nonostante il tumore polmonare sia associato con l'esposizione alle polveri sottili (inferiori ai 2.5  $\mu\text{m}$  di diametro), la mortalità non diminuisce in caso di riduzione dell'emissione delle polveri. (*Jemal A. et al., 2007*). Circa il 20% di tutti i tumori del polmone sono causati dalla combinazione di fattori

ambientali (asbesto e radon) e/o fattori genetici. Nonostante la maggioranza dei pazienti affetti da tumore del polmone siano fumatori, solo una minoranza dei forti fumatori (il 15% ) sviluppa nel corso della propria vita tale patologia, suggerendo che l'iniziazione e la progressione della patologia sia guidata dalla concomitante azione di fattori ambientali e fattori genetici. Si ritiene inoltre che anche fattori ormonali e virali, come il papilloma virus, possano essere coinvolti in tale processo (*Herbst et al., 2008*). I tumori polmonari vengono in genere suddivisi in due classi, small cell lung cancer (SCLC) o tumore del polmone a piccole cellule e non-small cell lung cancer (NSCLC) o tumore del polmone non a piccole cellule. Approssimativamente l'80-85% dei tumori del polmone sono NSCLC e includono i carcinomi a cellule squamose, gli adenocarcinomi e i carcinomi a grandi cellule (*Govindan et al., 2006*). Molti tumori del polmone correlati al fumo di sigaretta si sviluppano spesso nell'area centrale del polmone (solitamente carcinomi a cellule squamose e tumori polmonari a piccole cellule). Mentre è stato osservato che gli adenocarcinomi si sviluppano principalmente nelle zone periferiche (*Herbst et al., 2008*).

L'adenocarcinoma è attualmente il tumore polmonare più diffuso nell'Asia dell'Est, negli Stati Uniti e in molti paesi europei sta ormai superando il carcinoma a cellule squamose. Sebbene si sviluppi principalmente in fumatori e in ex fumatori, esso insorge in pazienti non fumatori più frequentemente di altri istotipi. L'adenocarcinoma polmonare è pertanto l'istotipo di tumore polmonare osservato più frequentemente in assenza di esposizione al fumo di sigaretta, a volte anche in soggetti molto giovani (*Stahel, 2007*). La sua incidenza sembra essere in crescita nel mondo, probabilmente sia in relazione al fatto che le moderne sigarette contengono una concentrazione maggiore di sostanze cancerogene, come le nitrosamine (*Hoffmann D et al., 1996*) sia in relazione al nuovo tipo di sigarette commercializzate. Negli ultimi anni infatti le sigarette vengono progettate con una minor concentrazione di nicotina allo scopo di ridurre l'abitudine al fumo. Tuttavia è stato

dimostrato che gli individui che fumano tali sigarette hanno modificato il proprio modo di fumare, aumentando l'intensità e la durata dell'inalazione, per compensare tali differenze (Gower S., and Hammond D., 2007).

Studi epidemiologici hanno indicato che fattori di predisposizione genetica possono avere un ruolo, assieme ai fattori ambientali, nello sviluppo dei tumori polmonari (Ahlbom et al., 1997; Sellers et al., 1992; Wood et al., 2000). Nei non fumatori il cancro polmonare potrebbe, pertanto, essere causato dalla prevalenza di fattori genetici di suscettibilità alla malattia e/o diminuzione di fattori di resistenza. Negli ultimi anni numerosi studi epidemiologici e previsioni di modelli genetici associati a modelli sperimentali murini fanno presupporre che diversi fattori genetici possono svolgere un ruolo fondamentale nello sviluppo di tale patologia. Lo studio di tale problema in modelli murini, mediante analisi di linkage genetico, ha dimostrato l'esistenza di molteplici loci associati con la predisposizione alla suscettibilità e con la resistenza ai tumori polmonari. Questa è definita come ereditarietà di tipo poligenico. Per analogia con il modello sperimentale murino, e in accordo con le previsioni di modelli genetici applicati ai risultati epidemiologici, ci si attende che il rischio di sviluppare un tumore polmonare nell'uomo sia determinato dal coinvolgimento di diversi loci genetici, almeno 5-10 (Peto J. and Houlston RS., 2001; Dragani TA, 2003).

Recentemente alcuni studi di associazione hanno tentato di identificare i loci genetici coinvolti nella suscettibilità al tumore del polmone (Risch A et Plass C., 2008). È stato individuato un locus sul cromosoma 15q25 che sembrerebbe fortemente associato con tale rischio (Hung RJ et al., 2008). Tuttavia alcuni studi ipotizzano che l'effetto del locus 15q25 sul rischio di sviluppare un tumore polmonare sia dovuto interamente alla sua correlazione con il principale fattore di rischio ambientale, il fumo di sigaretta (Galvan A. et al., 2010) vista la presenza sullo stesso locus del gene che codifica per il recettore

nicotinico dell'acetilcolina (*Freedman et al., 2008*). Alcuni studi di associazione hanno proposto altri due loci di suscettibilità sui cromosomi 5p15.33 (*Wang et al., 2008; McKay et al., 2008*) e 6p21 (*Wang et al., 2008*). Il maggior locus di suscettibilità per il tumore del polmone è stato individuato nel cromosoma 6, più precisamente nella regione 6q23-25 (*Bailey-Wilson et al., 2004; You et al., 2009*). Recenti studi hanno confermato, mediante analisi di linkage su famiglie, che la regione cromosomica 6q sarebbe coinvolta nel rischio di sviluppare tumore polmonare. Tale rischio sarebbe inoltre maggiore in individui non fumatori, o deboli fumatori, avvalorando pertanto l'ipotesi che i fattori genetici abbiano un ruolo di rilievo nella suscettibilità al cancro del polmone. Rimane tuttavia da chiarire se e come i geni che mappano in questa regione siano correlati allo sviluppo del tumore (*Amos et al., 2010*).

Studi molecolari citogenetici hanno mostrato che nella maggior parte dei tumori del polmone si verificano aberrazioni cromosomiche nel braccio corto del cromosoma 5 (*Ried et al., 1994; Balsara et al., 1997; Ullmann et al., 1998; Luk et al., 2001*).

Una diagnosi precoce del tumore del polmone potrebbe cambiare l'esito della patologia, con un miglioramento della prognosi. Tuttavia gli attuali strumenti sono molto costosi o non sufficientemente sensibili da consentire una diagnosi precoce. Attualmente le speranze sono rivolte verso i progressi tecnologici nell'oncologia molecolare che per la prima volta siano in grado combinare la genomica e l'epigenetica e prevedere tramite il loro ausilio dei programmi di screening epidemiologici (*Risch A. and Plass C., 2008*). I marcatori genetici potrebbero diventare infatti dei marcatori tumorali predittivi del rischio di sviluppare tumore polmonare, della prognosi o della sensibilità alla terapia utilizzata (*Herbst et al., 2008*).



## ***Scopo della Tesi***

Lo scopo del progetto del mio dottorato è stato l'identificazione di polimorfismi genetici associati con il rischio di sviluppare tumori polmonari nella popolazione sarda, che a causa dell'isolamento geografico mostra delle caratteristiche di omogeneità genetica che la rendono particolarmente adatta a questi tipi di studi. L'individuazione delle varianti genetiche associate con il rischio di sviluppare tali tumori permetterebbe l'identificazione, nella popolazione generale, degli individui ad alto rischio di sviluppare tale patologia e prevedere quindi per essi opportuni programmi di controllo che possano consentire una diagnosi precoce e quindi un miglioramento della prognosi.

## ***Materiali e Metodi***

## ***Casistica***

### ***Pazienti affetti da tumore polmonare***

I pazienti affetti da adenocarcinoma polmonare primitivo sono stati reclutati presso la Divisione di Chirurgia Toracica del P. O. Binaghi e il Dipartimento di Chirurgia Generale dell'Ospedale Oncologico Businco di Cagliari (2006-2009). Tutti i pazienti sono stati intervistati e informati circa lo scopo e le modalità di svolgimento del progetto, hanno inoltre firmato un consenso informato approvato dal Comitato Etico degli Ospedali coinvolti. Requisito fondamentale per la partecipazione al progetto è stata l'origine sarda da generazioni. Ai pazienti è stato prelevato un campione di sangue venoso periferico (8 mL) che è stato conservato a +4 °C per un periodo di tempo non superiore alle 3 settimane e successivamente processato per l'estrazione di DNA genomico. Inoltre, a tutti i pazienti sottoposti ad intervento chirurgico di resezione del tumore polmonare, è stato prelevato una porzione di tessuto neoplastico e una porzione di tessuto normale adiacente senza infiltrazioni neoplastiche. Il tessuto polmonare è stato conservato in una soluzione di RNALater (Ambion) per impedirne la degradazione e congelato a -80°C. Successivamente è stato utilizzato per l'estrazione di DNA e RNA.

### ***Controlli sani di popolazione***

La popolazione di controllo è composta da individui sani donatori di sangue presso il Centro di Trapianti d'Organo dell'Ospedale Brotzu di Cagliari e da individui sani partecipanti ad uno studio sui fattori di rischio cardiovascolari che volontariamente hanno deciso di partecipare a questo progetto. Così come i pazienti anche i controlli, sardi da generazioni, sono stati intervistati e informati sul progetto. Hanno inoltre firmato il consenso informato approvato dal Comitato Etico degli Ospedali coinvolti. A tali individui

è stata prelevata un aliquota di sangue venoso periferico (8 mL) che è stato utilizzato per l'estrazione di DNA genomico. I dati epidemiologici relativi sono stati raccolti in un database continuamente aggiornato.

## ***Metodiche***

### ***Estrazione di DNA genomico da sangue mediante PAXgene Blood DNA kit (PreAnalytix, QIAGEN)***

Tutti i partecipanti allo studio sono stati sottoposti al prelievo di circa 8 ml di sangue periferico mediante apposite provette vacutainer fornite dal kit e contenenti una soluzione di anticoagulazione e stabilizzazione del DNA che consente di conservarlo perfettamente a temperatura ambiente per 15 giorni. Il sangue contenuto in queste provette è stato trasferito per l'estrazione, in provette da 50 mL contenenti 25 mL di buffer BG1 (fornite direttamente dal kit). Queste sono state poi sottoposte a centrifugazione a 3500g per 3 minuti in maniera tale da consentire la formazione di un pellet di linfociti. Il surnattante viene eliminato e il pellet, che contiene ancora qualche eritrocita, viene lavato con 5 mL di buffer BG2 e successivamente centrifugato a 3500 g per 1 minuto. Il surnattante viene eliminato e il pellet di linfociti viene trattato con 5 ml di una soluzione di lisi (buffer BG3) e 25 µL di proteinasi K in bagnetto termostato a 65°C per 15 minuti. Una volta effettuata la lisi, al campione è stato aggiunto isopropanolo (5 mL) per consentire la formazione di un precipitato bianco di DNA. Dopo aver isolato il pellet per centrifugazione (3500 g per 1 minuto) e lavato con etanolo al 75% (V/V), viene fatto asciugare all'aria per consentire la completa eliminazione dell'etanolo che potrebbe interferire con le reazioni successive (ad es PCR). Il campione è stato poi risospeso in acqua sterile (1 mL) e conservato a -20°C fino al successivo utilizzo.

### ***Estrazione di DNA genomico da tessuto mediante QIAamp DNA Mini Kit (QIAGEN)***

A 20 mg di tessuto polmonare normale vengono aggiunti 180 µL di buffer ATL e 20 µL di Proteinasi K. Il campione è stato poi incubato a 56°C per minimo 3 ore e massimo overnight in un bagnetto termostato. Successivamente sono stati aggiunti 200 µL di buffer AL e il campione è stato posto per 10 minuti in bagnetto termostato a 70°C. In seguito sono stati aggiunti 200 µL di etanolo (96%) e il campione è stato trasferito in apposite colonne. Queste ultime sono state sottoposte a centrifugazione a 8000 g per 1 minuto e il DNA, che resta adsorbito sulla resina della colonna è stato lavato con il buffer AW1 e successivamente con il buffer AW2. La risospensione del DNA è stata effettuata con acqua sterile, mediante incubazione per 5 minuti e successiva eluizione per centrifugazione (8000 g per 1 minuto). Il campione di DNA così ottenuto è stato conservato a -20°C. Mediante tale metodica sono stati estratti i DNA di 5 pazienti di cui 2 affetti da adenocarcinoma polmonare e 3 affetti da carcinoma squamo cellulare.

### ***Estrazione di RNA totale da tessuto mediante Rneasy Mini Kit (QIAGEN)***

L'RNA totale è stato estratto da porzioni di tessuto polmonare normale e tumorale conservate a -80°C. Da ciascuna porzione di tessuto è stato prelevato un frammento del peso di circa 30 mg che è stato poi omogenizzato (con un omogenizzatore elettrico Turrax) in 600 µL di buffer RTL. I campioni sono stati quindi sottoposti a centrifugazione per 3 minuti a 14000 rpm a temperatura ambiente allo scopo di eliminare il materiale insolubile. Il surnattante è stato quindi prelevato e posto in una provetta da 1,5 mL. A questo sono stati aggiunti 600 µL di una soluzione di etanolo al 70% e la soluzione così ottenuta è stata sottoposta ad agitazione e successivamente dispensata all'interno della colonna posta in

una provetta da 2 mL. Tale colonna è stata sottoposta a centrifugazione per 15 secondi a 10000 rpm (a temperatura ambiente) e l'eluato è stato quindi eliminato. La colonna è stata poi sottoposta a diversi lavaggi prima con 700  $\mu$ L di buffer RW1 e poi con 500  $\mu$ L di buffer RPE (quest'ultimo per due volte), ciascuno dei quali seguito da centrifugazione per 15 secondi a 10000 rpm con eliminazione dell'eluato. La colonnina è stata trasferita in una nuova provetta da 1,5 mL e a questa sono stati aggiunti 50  $\mu$ L di acqua sterile (priva di DNasi e RNasi) per la risospensione dell'RNA. Dopo incubazione per 1 minuto circa, la colonna è stata sottoposta a centrifugazione e l'eluato di RNA così ottenuto è stato conservato a  $-20^{\circ}\text{C}$ .

### ***Valutazione qualitativa e quantitativa degli acidi nucleici***

La quantità degli acidi nucleici estratti è stata determinata spettrofotometricamente e ne è stato valutato il rapporto A260/A280. Per il DNA questo rapporto è sempre stato compreso tra 1,7 e 1,9; per l'RNA il rapporto è sempre stato compreso tra 1,8 e 2,0. Per i DNA utilizzati per la preparazione dei pool di DNA la quantizzazione è stata effettuata in triplicato mediante analisi fluorimetrica con il PicoGreen dsDNA Quantitation Kit (Invitrogen), una molecola in grado di emettere fluorescenza se intercalata in un doppio filamento, che consente quindi l'analisi di DNA a doppio filamento. Tale quantizzazione si è resa necessaria per poter assicurare la buona riuscita della preparazione dei pool e dell'analisi con gli SNP-array. Tutti gli altri campioni di acidi nucleici estratti sono stati quantizzati con il Nanodrop. Tutti i DNA sono stati sottoposti a corsa elettroforetica su gel di agarosio all'1% e colorati con bromuro di etidio per verificarne l'integrità. Solo quelli che hanno mostrato una media di frammenti maggiore di 18Kb sono stati selezionati per essere usati nei pool.



### ***Creazione dei pool di DNA***

Per la realizzazione dei pool abbiamo selezionato 39 pazienti affetti da adenocarcinoma polmonare e 39 individui sani di controllo scelti cercando di ridurre al minimo le differenze fenotipiche tra la popolazione dei casi e quella dei controlli. In particolare sono stati selezionati i campioni di DNA di pazienti in cui il tumore si è manifestato ad un'età inferiore ai 70 anni e i controlli sono stati selezionati per avere lo stesso genere, età uguale o maggiore e la stessa abitudine al fumo dei pazienti (Tabella 1). Requisito fondamentale è stata l'origine sarda dei pazienti e degli individui di controllo.

I DNA usati per la costruzione dei pool sono stati accuratamente quantizzati singolarmente mediante analisi fluorimetrica. Tutti i campioni utilizzati per la preparazione dei pool di DNA (39 DNA di pazienti affetti da adenocarcinoma polmonare e 39 DNA di controlli sani di popolazione) sono stati diluiti 1:100 e 1:200 in TE 1X (10mM Tris HCl e 1mM EDTA pH 7,5) in un volume finale di 100  $\mu$ L in una piastra da 96 pozzetti (Corning 96 Well Clear Flat Bottom Microplate). Contestualmente è stata preparata la diluizione dello standard (Lambda DNA 100  $\mu$ g/mL) per la costruzione della curva standard. A tutti i campioni così preparati e allo standard sono stati aggiunti 100  $\mu$ L della soluzione contenente i Picogreen (precedentemente diluito 1:200 in TE1X). La piastra è stata poi sottoposta ad analisi fluorimetrica (Tecan Ultra). I valori di fluorescenza dello standard sono stati utilizzati per la costruzione di una retta standard dalla quale ricavare i valori di concentrazione dei campioni in esame. Una volta effettuato il calcolo delle concentrazioni, tenendo conto delle rispettive diluizioni, la media di tali valori è stata usata per calcolare la quantità di materiale da prelevare ( $\mu$ L) per ciascun campione per ottenere le stesse quantità di DNA (200 ng). I vari pool sono stati quindi creati prelevando per ciascun campione la stessa quantità di acido nucleico (200 ng) e unendoli in un'unica provetta.

<i>Caratteristiche</i>		<i>Casi</i>	<i>Controlli</i>
<i>Appartenenza geografica</i>	Sardi	39	39
<i>Genere</i>	Maschi	26	26
	Femmine	13	13
<i>Età (mediana)</i>	-	59 (42-72)	63 (42-83)
<i>Abitudine al fumo</i>	Fumatori	31	31
	Non Fumatori	8	8
<i>Istotipo</i>	Adenocarcinoma	39	-

Tabella 1- Caratteristiche dei pazienti affetti da tumore polmonare e degli individui sani di controllo.

### ***Analisi Genome-wide sui pool***

I pool di DNA dei casi e dei controlli sono stati sottoposti ad analisi di circa 1,2 milioni di marcatori (per la maggior parte polimorfismi a singolo nucleotide, SNPs) dispersi su tutto il genoma mediante l'utilizzo del chip Human1M-Duo DNA Analysis BeadChip (Illumina) e della piattaforma Illumina BeadStation 500 System (Illumina). Tale chip contiene più di un milione di marcatori dei quali circa 950000 tag-SNPs e 100000 SNPs non presenti nelle mappe costruite dall'HapMap. Di questi polimorfismi 565718 si trovano all'interno di geni. Il chip contiene inoltre circa 260000 marcatori localizzati nelle regioni CNV (copy number variation), già note e in quelle recentemente individuate. Per prepararli per la successiva amplificazione 400 ng di DNA di ciascun pool sono stati denaturati mediante aggiunta di 8 µL di una soluzione di NaOH 0,1N. La reazione è stata lasciata incubare 10 minuti a temperatura ambiente. Sono poi stati aggiunti 135 µL di buffer MP1 (Neutralization Solution) per neutralizzare la soluzione basica di NaOH. Ad ogni campione sono stati aggiunti 150 µL della soluzione AMM (Amplification Master Mix) e sono stati lasciati amplificare per 22 ore a 37°C. La reazione di amplificazione del DNA genomico consente di incrementare uniformemente la quantità di DNA di alcune migliaia di volte. Successivamente il DNA è stato frammentato mediante l'incubazione di 1 ora a

37°C con 50 µL di FRG (Fragmentation Solution) e fatto precipitare attraverso l'incubazione di 5 minuti a 37°C con 100 µL del buffer PA1 (Precipitation Solution) e l'incubazione per 30 minuti a 4°C con aggiunta di 300 µL di isopropanolo seguito da centrifugazione a 4000 x g per 20 minuti a 4°C. Il DNA precipitato è stato risospeso in 46 µL di RA1 (Resuspension and Hybridization solution) mediante incubazione per 1 ora a 48°C. Al termine è avvenuta l'ibridizzazione del DNA sul chip mediante incubazione a 45°C per 20 ore. I chip sono poi stati sottoposti a vari lavaggi con buffer PB1 (per eliminare il DNA in eccesso o che non si è appaiato in maniera specifica) e a diverse reazioni (estensione di una singola base e colorazione per l'identificazione del nucleotide incorporato) attraverso l'uso del buffer RA1, una soluzione di formamide al 95% in EDTA 1mM e i buffer PB1, XC1, XC2, XC3, TEM, STM, ATM seguendo le istruzioni della casa produttrice. Infine i chip sono stati sottoposti a scansione mediante BeadArray Reader (Illumina). Entrambi i pool sono stati analizzati in triplicato.

I dati di output dello scanner Illumina sono rappresentati da intensità di fluorescenza; per ogni SNP sono presenti due segnali di intensità fluorescenza (Xraw, rosso e Yraw, verde) uno per ciascuno degli alleli del polimorfismo. Per ciascun marcatore presente sul chip abbiamo calcolato le frequenze alleliche, sia nel pool dei casi che in quello dei controlli, mediante la formula:

$$\text{frequenza dell' allele } X = \text{segnale raw } X / (\text{segnale raw } X + \text{segnale raw } Y)$$

## ***Scelta dei marcatori e disegno dei primer***

La scelta dei marcatori da sottoporre a successiva analisi si è basata sui risultati ottenuti dall'analisi delle differenze di frequenze alleliche tra il pool dei casi e quello dei controlli. Per ogni marcatore selezionato abbiamo quindi disegnato i primer (indicati in tabella 2) per l'amplificazione e la successiva analisi mediante pyrosequencing utilizzando un software apposito, Pyrosequencing<sup>TM</sup> Assay Design Software Version 1.0.

SNP	PRIMER FORWARD	PRIMER REVERSE	PRIMER SEQUENZA
rs4848272	biot-CCCACTTATGCTTTTACAAACAGG	ATCTATTAAGGGCATGAAGGCA	CAGTTTATATCTAAAAGCCA
rs787980	GGTCTATGTTCTTGGCTTCTGT	biotCAATTCCACCTCTGATTAACTCC	AGTGAATTGGCCTCAG
rs3792159	CCCCAAAAAGAAAAAACAGGA	biotTTTGACCCTTGCCTTTTGT	TACTTACTGCACATTGTAAC
rs6720502	biotAGTCAAGGGGTGTCTCTTAAAGTG	GATTTTGCCTAGAACCCATTATC	GCTTGCTTTTCTGTAACATA
rs724897	ACGTCCCAAATAAGTACCCTCTTC	biotCCACTACACCAGCTAATTTTG	TTTCTGTTCTGTGTATCTT
rs178117	CACAGGGATGGATAAAGTAGAGGT	biotGGCTATGAAGTATTGGGCAAGT	GATTGATCACACAGTCCTAT
rs5965727	biotAATATTAAGGTTCGCCAAATAGC	CAATCACGTGGGTTTTTGT	ATCATGGTAATGCTGG
rs 30616	biotCAAATTGGGGGACATTTAGTTGTA	CTCAAAGTGTGACGGAATCCA	CCAGTCTGTCTTTTCTGC

Tab.2- Primer utilizzati per l'analisi in PCR per amplificare le sequenze contenenti i polimorfismi in analisi e per il saggio al Pyrosequencing.

## ***PCR (Polymerase-Chain-Reaction) e analisi di SNPs mediante Pyrosequencing***

I campioni di DNA sono stati sottoposti ad amplificazioni mediante PCR utilizzando le opportune coppie di primer di cui uno presenta il 5' marcato con biotina. In generale, la miscela di reazione conteneva 50 ng di DNA, 0.2 mM di ciascun dNTPs, 1,5 mM di

MgCl<sub>2</sub>, 0,2uM della coppia di primer e 1.0U di Platinum Taq DNA Polymerase (Invitrogen). La reazione è stata effettuata in un termociclatore GeneAmp PCR System 9700 (Applied Biosystems) secondo i seguenti parametri per 40 cicli: 94°C per 30 secondi, 55-58°C per 30 secondi (varia a seconda della coppia di primer utilizzata) e 72°C per 1 minuto. Al termine della reazione, 3 µl di prodotto di PCR sono stati sottoposti a corsa elettroforetica su gel di agarosio al 2,5% per valutare la buona riuscita dell'amplificazione. I prodotti di PCR sono stati successivamente sottoposti a successive procedure per isolamento del frammento ottenuto dal primer biotinilato mediante l'utilizzo di un apposito sistema e di opportuni reagenti. Il frammento di PCR biotinilato è stato analizzato al Pyrosequencing (PSQ96MA, Biotage, AB, Uppsala, Sweden) mediante l'uso di un opportuno primer di sequenza, di enzimi (DNA Polimerasi, Apirasi, Sulfonilasi e Luciferasi), substrati (adenosina 5'fosfosolfato e luciferina) e dNTPs. In questo modo, durante la reazione di sintesi del filamento complementare a quello biotinilato mediante l'uso di un opportuno primer (primer di sequenza), di una DNA polimerasi e dell'aggiunta dei singoli dNTPs, si produce pirofosfato che reagendo con l'adenosina 5'fosfosolfato porta alla formazione di ATP. Quest'ultima guida la reazione che converte la luciferina in ossiluciferina che emette un segnale luminoso che viene registrato e visualizzato in un grafico dallo strumento. La quantità di segnale generato è direttamente proporzionale alla quantità di ATP prodotto, a sua volta proporzionale alla quantità di pirofosfato e quindi al numero di nucleotidi incorporati durante la reazione di sintesi. In questo modo è possibile ottenere sia il genotipo dei singoli campioni sia un'accurata stima delle frequenze alleliche di un determinato SNP in un pool di campioni.

### ***RT-PCR (Reverse-Transcription-Polymerase-Chain-Reaction)***

L'RNA è stato retrotrascritto in cDNA mediante l'utilizzo del Kit High Capacity cDNA Reverse Transcription (Applied Biosystem).

1,5µg di RNA sono stati diluiti in 10 µL di acqua. Ad ogni campione sono poi stati aggiunti 2µL di RT Buffer (10X), 2µL di Random primer (10X), 0,8µL di dNTPs (100mM), 1µL dell'enzima Multiscribe Reverse Transcriptase, 1µL di RNase Inhibitor e l'appropriata quantità d'acqua fino ad arrivare a 20 µL. Questa soluzione è stata poi incubata a 25°C per 10 minuti, 120 minuti a 55°C e la reazione di sintesi del filamento di cDNA è stata interrotta mediante incubazione per 5 secondi a 85°C. I campioni sono stati conservati a -20°C fino al successivo utilizzo.

### ***Analisi di espressione dei geni TRIO e SNX9 mediante Real-Time PCR***

I campioni di RNA sono stati selezionati tra quelli per cui era disponibile sia il prelievo di sangue, sia la porzione di tessuto polmonare normale. Su questi campioni è stata prima condotta un'analisi dei polimorfismi rs30616 per il gene TRIO e rs724897 per il gene SNX9 sul DNA genomico. allo scopo di valutarne i genotipi. La reazione di amplificazione (in un volume di 10 µl) è stata effettuata utilizzando 4µl di cDNA (2,5 ng/µL ottenuto dai singoli campioni di tessuto polmonare tumorale e normale, 5µl di Taq Man PCR Master Mix (Applied Biosystems) e 1 µl di sonde Taq Man specifiche che amplificano una porzione dell'mRNA dei geni TRIO (Hs 01125855\_m1) e SNX9 (Hs 00212006\_m1) su un termociclatore ABI PRISM 7300HT (Applied Biosystems). I parametri utilizzati sono costituiti da uno step iniziale di attivazione 95°C per 10 minuti, mentre i 40 cicli di amplificazione sono stati condotti a 95°C per 15 secondi e 60°C per 1 minuto. I campioni sono stati analizzati in triplicato ed è stato amplificato, oltre al gene target, anche un gene

housekeeping ( $\beta$ -actina; 4352935E Applied Biosystems) per valutare eventuali variazioni nella quantità di cDNA. Le differenze relative tra i campioni con i diversi genotipi sono state valutate con il metodo del  $\Delta\Delta C_t$  usando come calibratore il cDNA dell'RNA ottenuto da tessuto normale.

### ***Analisi statistiche***

I valori di frequenza allelica ottenuti dall'analisi dei segnali di fluorescenza degli arrays sono stati utilizzati per ricostruire il numero di cromosomi portante ciascuno uno dei due possibili alleli. L'equilibrio di Hardy-Weinberg è stato valutato mediante il test del  $\chi^2$  (chi-quadro) in una tavola di contingenza di frequenze genotipiche attese contro frequenze genotipiche osservate (Weir, 1996). Nell'analisi dei pool di DNA, la correlazione tra i dati di frequenze alleliche ottenute mediante microarray e quelle ottenute mediante Pyrosequencing sono state misurate attraverso il coefficiente di correlazione di Pearson. Le differenze in termini di frequenze alleliche tra il pool dei casi e quello dei controlli sono state valutate mediante il test del chi-quadro. Le OR (Odds Ratio) e i relativi intervalli di confidenza (95%) sono state calcolate con il metodo della regressione logistica allo scopo di valutare l'associazione tra lo stato di malattia e il genotipo.

## ***Risultati e Discussione***



### ***Analisi genome-wide mediante microarray***

Per l'individuazione delle varianti genetiche associate ad un aumentato rischio di sviluppare tumore polmonare abbiamo analizzato in triplicato più di 1,2 milioni di marcatori genetici dispersi sull'intero genoma su un pool costituito dal DNA di 39 pazienti affetti da adenocarcinoma polmonare e su un pool costituito dal DNA di 39 individui sani di controllo, selezionati per avere lo stesso genere, abitudine al fumo ed età uguale o maggiore rispetto ai pazienti. Le caratteristiche dei pazienti e dei relativi controlli sono riportate in tabella 1. Le analisi sull'intero genoma sono state condotte in triplicato per migliorare l'attendibilità dei risultati ottenuti, ciascuna replica dei casi è stata quindi analizzata sullo stesso chip Human 1M-Duo con una replica dei controlli. Per i due alleli di ogni polimorfismo presente nei chip dell'array è stata quindi calcolata la frequenza allelica per ciascuna replica sia nel pool dei casi che in quello dei controlli mediante la formula:

$$f(A)=X_{raw}/(X_{raw}+Y_{raw})$$

dove  $X_{raw}$  e  $Y_{raw}$  sono i segnali di intensità di fluorescenza ottenuti per i due alleli del polimorfismo. Per verificare la riproducibilità dei risultati ottenuti abbiamo applicato il coefficiente di correlazione di Pearson tra i valori di frequenza allelica dei triplicati tecnici e questo è risultato compreso nell'intervallo tra 0.9584 e 0.9732 evidenziando quindi una buona correlazione tra i risultati delle repliche. Una volta calcolata la media delle frequenze alleliche per i due pool, abbiamo scelto come strumento di selezione il coefficiente di variazione (rapporto tra la deviazione standard e la media) che ci ha consentito di escludere dall'analisi quei polimorfismi con un valore di frequenza allelica che si discostava di molto nelle repliche. Il valore del coefficiente di variazione imposto come limite accettabile è stato dello 0.05.

A partire dai valori di frequenza allelica abbiamo ricostruito il numero di cromosomi portanti ciascun allele nel pool (frequenza allelica \* 78) e abbiamo applicato per ciascun marcatore il test del chi-quadrato per identificare quei marcatori che presentavano differenze di frequenze alleliche statisticamente significative tra il pool dei casi e quello dei controlli. e abbiamo deciso di ritenere come statisticamente associati con il rischio di sviluppare tumore polmonare quei polimorfismi che mostravano una distribuzione del chi quadrato inferiore a 0.05 ( $p < 0.05$ ).

I marcatori che abbiamo selezionato per le successive analisi rispettano quindi i seguenti requisiti:

- significativa differenza di frequenza allelica tra il pool dei casi e dei controlli: distribuzione del chi quadrato inferiore a 0.05 ( $p < 0.05$ );
- coefficiente di variazione, espresso in percentuale, tra i replicati tecnici per il pool dei casi di tumore inferiore a 5 (C.V.  $< 5$ );
- coefficiente di variazione, espresso in percentuale, tra i replicati tecnici per il pool dei controlli inferiore a 5 (C.V.  $< 5$ ).

Abbiamo così individuato circa 10000 marcatori dispersi su tutto il genoma statisticamente associati con il rischio di sviluppare tumore polmonare.

Alcuni studi hanno ipotizzato che l'utilizzo dei pool di DNA non sia appropriato nelle analisi genome-wide perché non sarebbero in grado di produrre lo stesso risultato che si otterrebbe con l'analisi sui singoli campioni in seguito all'insorgenza di errori nell'analisi dei dati, che se non opportunamente corretti potrebbero far incorrere l'operatore in facili falsi positivi (*Barratt BJ et al., 2002*). Altri studi sostengono invece che tale metodica possa essere utilizzata per l'individuazione delle varianti genetiche responsabili dello sviluppo della patologie complesse. Consente infatti la genotipizzazione di un grande numero di individui con una notevole riduzione dei costi. L'approccio del DNA-pooling

può essere pertanto usato come primo step nello studio di patologie complesse (*Pearson JV. Et al, 2007; Zhang H et al., 2008; Macgregor et al., 2008; Sebastiani et al., 2008*).

Allo scopo di identificare i marcatori da sottoporre a successive analisi abbiamo osservato i polimorfismi rispetto alla loro posizione nel genoma, evidenziando quei polimorfismi che presentavano differenze di frequenze alleliche tra il pool dei casi e quello dei controlli statisticamente significative e quei polimorfismi che mostravano un coefficiente di variazione inferiore a 0.05. In tal modo è stato possibile individuare le regioni genomiche che presentavano un'elevata densità di polimorfismi (almeno 10) associati allo sviluppo del tumore polmonare, escludendo dalle successive analisi i polimorfismi che pur avendo i requisiti da noi richiesti erano isolati e pertanto potevano essere dei falsi positivi. Con tale approccio abbiamo individuato numerosi loci genetici possibilmente associati al rischio di sviluppare tumore polmonare e mediante l'uso del software BeadStudio Genome Viewer abbiamo individuato i geni localizzati nei loci genetici identificati, come mostrato nella tabella 3.

Tab.3-Localizzazione dei loci individuati e geni presenti al loro interno

CROMOSOMA 1	1p13.2	HIPK1
	1p22.3	COL24A1
	1p36.13	ARHGEF10L-RCC2
CROMOSOMA 2	2p14	WDR92-PNO1
	2p15	PPP3R1
	2q32.2	WDR75-KRT18P19
	2q33.1	HSPD1-HSPE1-COQ10B-RFTN2-MOBKL3-SF3B1- LOC729342
	2q35	XRCC5
	2q36.2	CUL3
	2q13	TMEM87B
	2q14.1	TFCP2L1-SCTR
	2q14- q32	COL5A2
CROMOSOMA 3	3p11.2	EPHA3
	3q23- q24	CLSTN2
	3q12	ABI3BP
CROMOSOMA 4	4q21.23	AGPAT9
	4p15	PCDH7
	4p15.1	LOC642305
	4p13	LOC402175-KCTD8
	4q21	FGF5-C4orf22
CROMOSOMA 5	5p15.2	TRIO
	5q15	LOC391813 -RIOK2
	5q21	MCC
	5q22.2	YTHDC2-TSSK1B
	5q31.3	KCTD16
	5q12	CD180
CROMOSOMA 6	6q14.2	PRSS35-SNAP91
	6q22-23	MYB
	6q25.1- q26	SNX9
	6q25.3	SYNJ2
	6p21	NOL5BP
	6p22.1	TRIM27-C4orf100-LOC401242
CROMOSOMA 7	7p13	GLI3
	7q21.2	ANKIB1
	7q31.2	TFEC
CROMOSOMA 12	12q24.31	MPHOSPH9
		SBNO1
		RILPL2
CROMOSOMA 17	17q21.3	SKAP1-HOXB1-CALCOCO2-SPOP-SLC35B1-FAM117A- LOC124865-COL1A1
CROMOSOMA 18	18q21.3	DCC
CROMOSOMA 22	22cen- q12.3	HPS4
CROMOSOMA X	Xq22.2	MCART6

Tra questi loci alcuni presentano un numero considerevole di geni di interessante rilievo. Nella loro analisi abbiamo posto la nostra attenzione in quei geni correlati con i meccanismi di mantenimento delle normali strutture cellulari, nella cancerogenesi a livello di qualsiasi organo o in patologie che riguardassero comunque alterazioni a carico dell'organo da noi studiato e quindi del polmone.

Abbiamo evidenziato un importante locus sul braccio lungo del cromosoma 2, dove è presente il gene XRCC5 già associato alla malattia ostruttiva polmonare cronica (COPD) (*Hersh CP et al., 2010*). Abbiamo individuato due loci sul cromosoma 5, di cui quello sul braccio corto è in prossimità di un altro locus già descritto come coinvolto nella suscettibilità al tumore del polmone (*Wang et al., 2008; McKay et al., 2008*). All'interno del locus da noi identificato sul 5p mappa il gene TRIO che sembra svolgere un ruolo importante nella leucemia, nel tumore polmonare a cellule squamose, nel glioblastoma e nel carcinoma a cellule squamose dell'esofago (*Zheng et al., 2004; Coe et al., 2005; Garnis et al., 2005; Salhia et al., 2008; Chattopadhyay et al., 2010*). Un altro locus, anch'esso molto esteso sul cromosoma 6, è già stato descritto come il precedente nella suscettibilità al tumore polmonare (*Bailey-Wilson, 2004; You, 2009; Wang et al., 2008*). Al suo interno abbiamo ritenuto di particolare interesse il gene SNX9, il cui prodotto proteico appartiene ad una famiglia di proteine coinvolte nei meccanismi di degradazione del recettore per il fattore di crescita dell'epidermide (EGFR) (*Lin, 2002*). Un altro locus di interesse è quello sul cromosoma 12, in cui mappa il gene SBNO1, per il quale è stato descritto il coinvolgimento nel tumore polmonare NSCLC (*Suzuki C. et al., 2007*). Nel locus sul cromosoma 17 mappano 2 geni coinvolti nella morfogenesi cellulare: HOXB1, codificante per fattori di trascrizione coinvolti appunto nella morfogenesi cellulare e COL1A, codificante per il collagene e associato all'osteoporosi (*Jin H et al., 2010*). Nel locus presente sul cromosoma 22 abbiamo riscontrato un'elevata densità di polimorfismi

con differenze di frequenze alleliche statisticamente significative tra i casi e i controlli all'interno del gene DCC, noto oncosoppressore coinvolto nel tumore dell'esofago e del colon (*Hibi K et al., 2009*) e nel gene HPS4, le cui alterazioni vengono descritte nella fibrosi polmonare (*Anderson PD et al., 2003*). Infine un locus sul cromosoma X, in prossimità del gene MCART6, un trasportatore mitocondriale per il quale non è stato descritto alcun coinvolgimento in nessuna patologia.

Prima di procedere alle analisi successive, abbiamo voluto analizzare le frequenze alleliche nei due pool attraverso l'utilizzo di una metodica più robusta rispetto a quella degli array per la determinazione della frequenza allelica. A tale scopo abbiamo selezionato 8 polimorfismi (appartenenti a 6 dei loci genetici identificati) risultanti possedere differenze di frequenze alleliche statisticamente significative tra il pool dei casi e quello dei controlli e ne abbiamo valutato la frequenza allelica tramite Pyrosequencing sugli stessi pool precedentemente utilizzati. I risultati di tali analisi sono riportate in tabella 4.

SNP	Tipo polimorfismo	Chr	Posizione (Mb)	Frequenze alleliche Array		Frequenze alleliche Pyrosequencing		p array	p pyro
				ADCA	Controlli	ADCA	Contr olli		
rs4848272	C/T	2	112.57	0,308	0,625	0,363	0,658	6.04E-05	0.0002298
rs787980	A/G	2	197.25	0,890	0,595	0,941	0,696	2.87E-05	9.22E-05
rs3792159	A/G	2	198.11	0,916	0,655	0,937	0,719	0.0001053	0.0003216
rs6720502	C/T	2	225.09	0,785	0,482	0,894	0,654	0.0001315	0.0002654
rs30616	C/T	5	14.3	0,781	0,484	0,913	0,663	0.0001315	0.0001959
rs724897	G/T	6	158.28	0,771	0,529	0,891	0,735	0.0014509	0.014801
rs178117	A/G	X	103.23	0,564	0,872	0,676	0,920	1.96E-05	0.0001379
rs5965727	A/G	X	145.21	0,657	0,272	0,648	0,254	1.45E-06	6.23E-07

Tab. 4- Confronto tra i risultati di frequenze alleliche ottenute mediante la metodica dell'array e quella del pyrosequencing per i polimorfismi selezionati. Le frequenze alleliche riportate si riferiscono agli alleli A e T degli SNPs.

Il coefficiente di correlazione di Pearson tra i dati degli array e quelli ottenuti mediante

pyrosequencing è 0,97 per gli ADCA e 0,90 per i controlli. Tali risultati indicano come vi sia una buona correlazione tra i dati di frequenze alleliche ottenute con le due metodiche.

Dai dati di frequenze alleliche ottenute mediante Pyrosequencing abbiamo ricostruito il numero di cromosomi portanti ciascuno uno dei due possibili alleli del polimorfismo per il pool dei casi e dei controlli e ne abbiamo valutato statisticamente le differenze. Tutti i polimorfismi analizzati mostrano significative differenze di frequenze alleliche tra i due pool con p variabile tra 0.0148 e  $6 \times 10^{-7}$ .

Allo scopo di valutare le reali frequenze alleliche dei due pool, abbiamo analizzato questi stessi marcatori sui singoli campioni di DNA dei pazienti e dei controlli facenti parte dei pool. I risultati riportati in tabella 5 mostrano come le frequenze alleliche reali ottenute dall'analisi dei genotipi sui singoli campioni correlano perfettamente con i risultati precedentemente ottenuti (i coefficienti di correlazione di Pearson tra i dati di frequenze alleliche dei pool ottenute mediante Pyrosequencing e quelle reali calcolate sulla base dei dati genotipici sono di 0,9897 per gli ADCA e 0,9867 per i controlli).

SNP	Frequenze alleliche reali (a)		Genotipi ADCA			Genotipi Controlli			OR <sup>2</sup>	95% CI (aa vs ab or bb)	p
	ADCA	Controlli	aa	ab	bb	aa <sup>1</sup>	ab	bb			
rs4848272	0,410	0,692	5	22	12	19	16	4	0.15	0.05-0.478	0.0006
rs787980	0,936	0,731	34	5	0	22	13	4	0.19	0.06-0.59	0.0025
rs3792159	0,949	0,782	35	4	0	24	13	2	0.18	0.05-0.62	0.0037
rs6720502	0,897	0,667	31	8	0	16	20	3	0.18	0.07-0.49	0.0005
rs 30616	0,961	0,724	35	3	0	21	13	4	0.1	0.03-0.39	0.0002
rs724897	0,936	0,769	34	5	0	24	12	3	0.24	0.08-0.74	0.0095
rs178117	0,654	0,882	23	5	11	33	1	4	4.59	1.47-14.31	0.006
rs5965727	0,295	0,705	9	5	25	26	3	10	6.67	2.45-18.11	0.0001

Tabella 5- Frequenze alleliche reali, OR e relativi intervalli di confidenza e significatività per i polimorfismi analizzati mediante metodica del Pyrosequencing.<sup>1</sup>Per tutti gli SNPs l'allele a rappresenta gli alleli A o T tranne che per il polimorfismo rs5965727 dove indica l'allele G. <sup>2</sup>l'OR è stata calcolata confrontando il numero degli individui portanti l'allele raro b rispetto al numero degli individui omozigoti per l'allele comune

Sebbene la dimensione dei campioni analizzati non sia numerosa è comunque possibile osservare come vi sia indicazione di un'associazione con lo sviluppo di adenocarcinoma polmonare per tutti i polimorfismi analizzati suggeriti dai valori di OR e relativi intervalli di confidenza (calcolati seguendo il modello dominante aa vs ab or bb). Quindi presumibilmente questi polimorfismi potrebbero essere coinvolti nel rischio di sviluppare tumore polmonare o essere in forte *linkage disequilibrium* con altri polimorfismi direttamente coinvolti nella suscettibilità genetica a tale patologia.

Tali dati confermano che le analisi sulle varianti genetiche disperse sull'intero genoma possano essere condotte non solo sui singoli campioni di DNA ma anche su pool costituiti da diversi campioni di DNA appartenenti e differenti pazienti, senza che questo possa incidere sull'efficacia dell'esperimento, così come alcuni lavori in letteratura indicano (Pearson JV. *Et al*, 2007; Steer *et al.*, 2007; Kirov *et al.*, 2008). Tra tutti i loci individuati abbiamo riposto la nostra attenzione su due in particolare che presentavano al loro interno il maggior numero di polimorfismi con differenze di frequenze alleliche statisticamente significative tra i casi e i controlli e che sono localizzati in prossimità di loci già descritti come associati allo sviluppo del tumore polmonare. Il primo locus si trova nel cromosoma 5 (5p15.2) mentre l'altro sul cromosoma 6 (6q25.1-q26).



## ***Il locus 5p15.2***

All'interno del cromosoma 5 abbiamo individuato una regione di circa 200 kb in cui sono presenti numerosi polimorfismi risultati con differenza di frequenza allelica statisticamente significativa tra il gruppo dei casi e quello dei controlli. Tali polimorfismi oltre ad avere una significatività indicata da un valore  $p < 0.05$ , come indicato in tabella 6, mostrano anche un coefficiente di variazione tra i replicati tecnici per il pool dei casi di tumore e dei controlli inferiore al 5%.

SNP	Posizione (in Mb)	Frequenza allele A		Coefficiente di variazione %		p
		Casi	Controlli	Casi	Controlli	
rs26136	14.270093	0.55	0.37	2.49	4.2	0.0266294
rs32390	14.272717	0.72	0.55	3.14	4.11	0.0334615
rs354289	14.290287	0.86	0.65	0.44	2.48	0.0022323
rs6876763	14.294854	0.65	0.46	1.74	1	0.0150993
rs1967180	14.298361	0.65	0.45	3.7	3.55	0.012882
rs165095	14.318371	0.65	0.48	4.8	0.79	0.0294743
<u>rs30616</u>	14.320173	0.78	0.48	2.85	0.56	8.39E-05
rs39856	14.322644	0.77	0.54	1.95	1.99	0.0027625
rs30619	14.330686	0.75	0.55	1.59	0.89	0.007959
rs17302408	14.336073	0.82	0.58	1.32	4.67	0.0009215
rs773702	14.337628	0.74	0.51	1.98	1.43	0.0037685
rs17304403	14.415334	0.56	0.35	1.82	3.45	0.0072174
rs434279	14.434816	0.63	0.46	2.17	3.2	0.0272707
rs378369	14.447611	0.62	0.45	3.67	2.86	0.0396766
rs30774	14.450238	0.6	0.44	1.34	3.94	0.0411442
rs26086	14.560905	0.81	0.65	1.96	0.79	0.0238874

Tab. 6 - Frequenze alleliche, coefficiente di variazione e significatività dei polimorfismi presenti nel locus 5p15.2 ottenuti mediante la metodica dell'array.

Studi presenti in letteratura hanno indicato la presenza di un locus di suscettibilità allo sviluppo del tumore polmonare sul cromosoma 5, in posizione 5p15.33 (*Wang et al., 2008; McKay et al., 2008*), quindi in una regione vicina a quella da noi individuata.

### ***Analisi del polimorfismo rs30616***

Il polimorfismo rs30616 che abbiamo analizzato sui singoli campioni di DNA costituenti i pool è del tipo C/T ed è localizzato in una regione intronica del gene TRIO.

Dalle analisi eseguite in questa prima fase abbiamo quindi ipotizzato che il polimorfismo rs30616 potesse rappresentare un marcatore in grado di indicare una maggiore suscettibilità allo sviluppo del tumore polmonare. Abbiamo quindi deciso di ampliare la casistica analizzando il medesimo polimorfismo su ulteriori 51 casi di adenocarcinoma polmonare e su 266 individui sani come controllo. Inoltre abbiamo deciso di analizzare il genotipo degli stessi polimorfismi su 35 carcinomi squamocellulari per ricercare l'esistenza di eventuali differenze genotipiche all'interno di una diversa tipologia di tumore polmonare.

I dati relativi ai nuovi campioni esaminati sono mostrati in tabella 7.

	<i>Genere</i>		<i>Età (mediana)</i>	<i>Abitudine al fumo</i>
	M	F		
<i>Adenocarcinomi</i>	34	19	60 (44-76)	50
<i>Carcinomi squamocellulari</i>	30	5	61 (42-74)	34
<i>Controlli</i>	222	53	42	143

Tabella 7– Caratteristiche dei casi di adenocarcinoma, carcinoma squamocellulare e dei controlli sottoposti a genotipizzazione mediante Pyrosequencing per i polimorfismi rs30616 e rs724897.

Così come già fatto in precedenza, abbiamo analizzato il genotipo mediante Pyrosequencing e calcolato le frequenze alleliche nei casi di tumore e nei controlli sull'intera casistica.

<i>Campione</i>	<i>Genotipi</i>			<i>Frequenza allele T</i>
	<i>T/T</i>	<i>T/C</i>	<i>C/C</i>	
<i>ADCA</i>	68	20	1	0.88
<i>Carcinomi squamocellulari</i>	25	9	1	0.84
<i>Controlli</i>	192	92	22	0.78

Tab. 8- Genotipi e frequenze alleliche dei campioni dell'intera casistica per il polimorfismo rs30616.

Dai risultati ottenuti, mostrati in tabella 8, possiamo osservare che abbiamo riscontrato un solo omozigote C/C tra i casi di adenocarcinoma analizzati. Abbiamo poi calcolato i valori di Odds Ratio per valutare l'associazione tra il polimorfismo analizzato e il rischio di sviluppare tumore polmonare.

<i>Modello</i>	<i>ADCA vs Controlli</i>		<i>Carcinomi squamo cellulari vs Controlli</i>	
	<i>Significatività valore-p</i>	<i>OR (CI 95%)</i>	<i>Significatività valore-p</i>	<i>OR (CI 95%)</i>
<i>TT vs TC o CC</i>	0.0168	0.52 (0.30-0.89)	0.3117	0.67 (0.31-1.45)
<i>TT vs TC</i>	0.0841	0.61 (0.35-1.07)	0.4832	0.75 (0.34-1.67)
<i>TT vs CC</i>	0.0196	0.13 (0.02-0.97)	0.0926	0.35 (0.05-2.70)

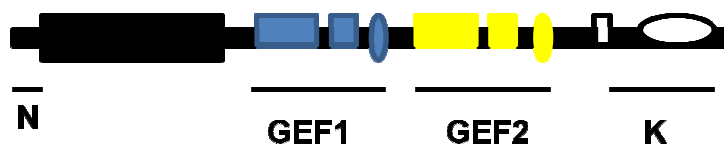
Tab.9- Significatività espressa tramite valore-p, valori di Odds Ratio e relativi Intervalli di Confidenza al 95% dei campioni dell'intera casistica per il polimorfismo rs30616.

Per il polimorfismo rs30616, in relazione al genotipo, abbiamo osservato una significativa associazione tra la presenza dell'allele raro in omozigosi e un ridotto rischio di sviluppare

adenocarcinoma polmonare rispetto a quanto osservato nei portatori dell'allele comune in omozigosi (OR uguale a 0.13 con CI 95% compreso tra 0.02 e 0.97). Inoltre, dall'analisi effettuata è risultato che i portatori della variante allelica C (omozigoti C/C e eterozigoti C/T) hanno un ridotto rischio di sviluppare adenocarcinoma polmonare (OR uguale a 0.52 con CI 95% compreso tra 0.30 e 0.89) rispetto agli omozigoti T/T. Non abbiamo riscontrato invece una significativa associazione per questo polimorfismo e il rischio di sviluppare carcinoma squamo cellulare tra i casi di tumore analizzati. Abbiamo infatti osservato un valore di OR uguale 0.67 con CI95% compreso tra 0.31 e 1.45. In base ai risultati ottenuti possiamo quindi affermare che il polimorfismo rs30616 sia associato con il rischio di sviluppare l'adenocarcinoma polmonare ma non sia invece associato con lo sviluppo di carcinoma squamocellulare.

### ***Il gene TRIO***

Il polimorfismo rs30616 si trova all'interno del gene TRIO (triple functional domain PTPRF interacting) localizzato in posizione 5p15.2. La proteina codificata da tale gene contiene tre domini funzionali: uno treonino-chinasico e altri due con attività GTPasica, chiamati GEFD1 e GEFD2, coinvolti nei meccanismi di trasduzione del segnale della famiglia di proteine Rho (*Garnis et al., 2005*).



lla

Trio agisce da modulatore citoscheletrico attivando le pathways di Rho e/o Rac. Esso fa infatti parte di un gruppo di proteine regolatrici chiamate GEFs (Guanine Nucleotide Exchange Factors) che funzionano da modulatori dell'attivazione/inattivazione del ciclo delle proteine con attività GTPasica della famiglia Rho. Attraverso il legame con la proteina inattiva, ne inducono l'attivazione tramite un cambio conformazionale che determina la liberazione di GDP (*Lane et al., 2008*). Aumentati valori dell'espressione proteica dei membri della famiglia delle proteine Rho sono stati riscontrati in alcuni tipi di tumore, tra i quali quello della mammella (*Jiang et al., 2003*). Lane e collaboratori hanno mostrato per la prima volta che, nel cancro della mammella, i livelli di Trio sono significativamente maggiori nel tessuto tumorale rispetto a quello normale, inoltre elevati livelli di Trio sono stati riscontrati nei tumori con una prognosi sfavorevole, suggerendone un possibile impiego come fattore prognostico in questo tipo di tumore (*Lane et al., 2008*). Inoltre è stato dimostrato che la proteina Trio abbia un ruolo fondamentale nell'orientamento degli assoni, nella migrazione dei neuroni e nella motilità cellulare (*Bateman J. and Vactor D., 2001*) così come nella regolazione dell'adesione cellulare (*Medley et al., 2003*). Già alcuni studi precedenti avevano correlato il gene TRIO alla tumorigenesi, è stato infatti proposto come oncogene nella leucemia (*Zheng et al., 2004*). Studi presenti in letteratura hanno mostrato l'amplificazione e overespressione del gene TRIO nel tumore polmonare a cellule squamose e una sua frequente amplificazione nelle lesioni pre-invasive suggerendone quindi non solo un suo possibile coinvolgimento nel tumore polmonare ma anche una sua implicazione nelle fasi precoci della tumorigenesi (*Garnis et al., 2005*). Inoltre è stato correlato alla tumorigenesi in linee cellulari di tumore polmonare a piccole cellule (SCLC) (*Coe et al., 2005*). Altri studi hanno indicato un'associazione tra l'amplificazione del gene TRIO e il carcinoma a cellule squamose dell'esofago (*Chattopadhyay et al., 2010*) e il tumore della vescica (*Zheng et al., 2004*).

Un' aumentata espressione del gene è stata riscontrata anche nel glioblastoma (*Salhia et al., 2008*). Alcuni studi hanno messo in luce, mediante la metodica del Comparative Genomic Hybridization (CGH), come l'amplificazione della regione 5p sia un evento comune nei NSCLC. Studi molecolari citogenetici hanno mostrato, inoltre, che nella maggior parte dei tumori del polmone si verificano aberrazioni cromosomiche nel braccio corto del cromosoma 5 (*Ried et al., 1994; Balsara et al., 1997; Ullmann et al., 1998; Luk et al., 2001*).

### ***Analisi dell'espressione genica***

La posizione e il tipo di polimorfismo è di grande importanza poiché può determinare l'effetto biologico del polimorfismo stesso. Polimorfismi non-sinonimi infatti determinerebbero un cambiamento nel codone e quindi nell'aminoacido codificato con la conseguente sintesi di una proteina alterata (*Bernig T. and Chanock S.J., 2006*). È altresì noto che la presenza di polimorfismi all'interno di regioni regolatorie può interessare l'espressione genica. Infatti a livello degli introni è possibile trovare alcune sequenze che funzionano da promotori per la trascrizione, oppure siti che regolano lo splicing alternativo. Una mutazione nelle sequenze che svolgono queste funzioni potrebbero influenzare l'espressione genica in maniera positiva, negativa o essere ininfluenta, esattamente come accade nelle sequenze codificanti degli esoni (*Morley et al., 2004*). Per approfondire il ruolo di questi polimorfismi nell'espressione genica, e cioè se fossero in grado di alterarla, abbiamo analizzato dal punto di vista quantitativo mediante PCR Real-Time, l'espressione del gene TRIO nel tessuto normale di pazienti affetti da adenocarcinoma polmonare facenti parti dei pool dei quali avevamo già analizzato il genotipo per il polimorfismo considerato.

### ***Analisi dell'espressione del gene TRIO***

Abbiamo voluto valutare le due varianti alleliche del polimorfismo potessero essere associate a eventuali differenze di espressione del gene TRIO. Abbiamo quindi analizzato, mediante Real-Time PCR l'espressione del gene in esame in 15 campioni di tessuto normale di polmone (elencati in tabella 10), di cui 6 con il genotipo T/C e 9 con il genotipo T/T per il polimorfismo rs30616.

<i>Campione</i>	<i>Sesso</i>	<i>Età</i>	<i>Abitudine al fumo</i>	<i>Genotipo rs30616</i>	<i>Genotipo rs724897</i>
LS16	M	49	si	T/C	T/T
LS21	F	42	si	T/T	G/T
LS29	M	56	si		G/T
LS54	F	63	si	T/T	T/T
LS57	M	58	si		G/T
LS68	F	54	si	T/T	T/T
LS96	F	65	si		G/T
LS109	F	57	si	T/T	T/T
LS116	M	60	si		
LS119	M	72	si	T/C	G/T
LS134	M	65	si	T/T	
LS161	F	73	si		G/T
LS166	M	68	si	T/T	T/T
LS177	F	56	si	T/T	T/T
LS186	M	66	si	T/T	
LS192	M	65	si	T/C	
LS210	M	58	si		G/T
LS286	M	66	si	T/C	
LD195	F	54	no	T/T	T/T
LD254	F	60	si	T/C	
LD270	M	56	si		G/T

Tabella 10 - Caratteristiche dei campioni sottoposti ad analisi di espressione genica in Real Time per i geni TRIO E SNX9



I risultati ottenuti sono riportati in figura 2.

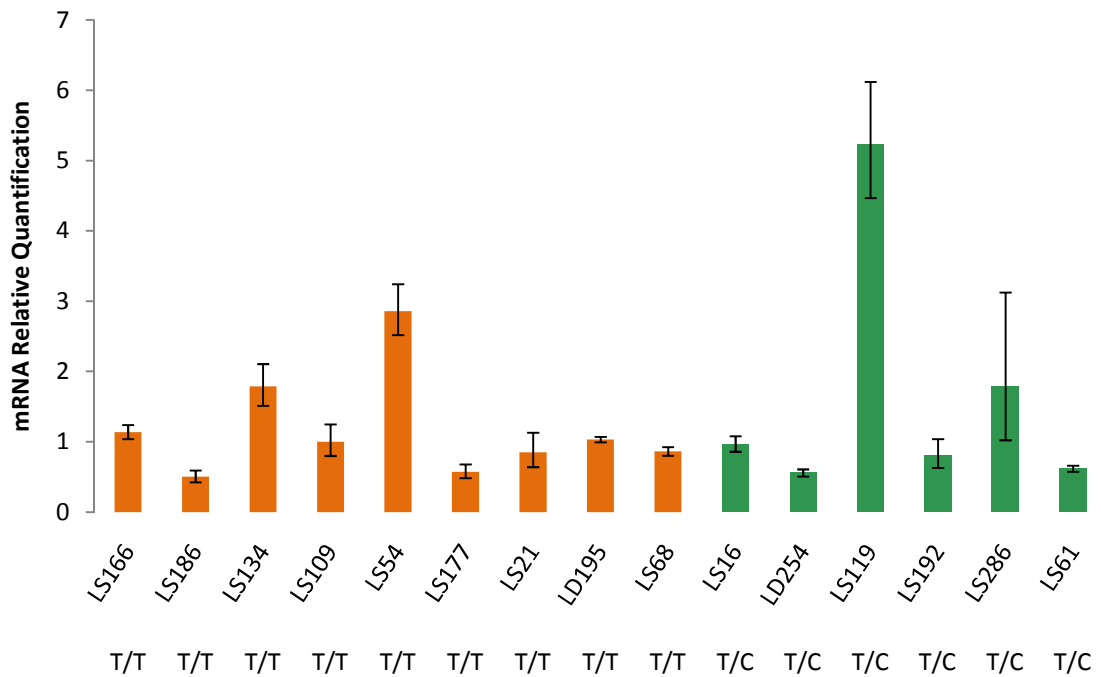


Fig. 2 - Analisi di espressione del gene TRIO in 15 campioni di tessuto normale con differente genotipo per il polimorfismo rs30616.

Mediante il test t abbiamo valutato se esistessero differenze di espressione tra i due gruppi portanti genotipo differente, ottenendo un valore di  $p=0.479$  quindi non significativo. Non possiamo quindi affermare che la presenza di una variante allelica per il polimorfismo rs30616 che avevamo individuato come associato con lo sviluppo di adenocarcinoma polmonare sia correlabile a differenze di espressione del gene TRIO.

### ***Il locus 6q25.1-q26***

Tra i loci individuati mediante analisi genome-wide per la ricerca delle varianti genetiche associate allo sviluppo del tumore polmonare abbiamo ritenuto di grande interesse un locus sul cromosoma 6 in posizione 6q25.1-q26. Tale regione presenta numerosi polimorfismi con differenza di frequenza allelica statisticamente significativa tra il gruppo dei casi e quello dei controlli. Infatti tali polimorfismi presentano, come mostrato in tabella 10, un valore di significatività statistica di  $p < 0.05$  e un coefficiente di variazione tra i replicati tecnici per il pool dei casi di tumore e quello dei controlli inferiore al 5%.

SNP	Posizione (in Mb)	Frequenza allele A		Coefficiente di variazione %		p
		Casi	Controlli	Casi	Controlli	
rs6924813	158.23712	0.75	0.52	2.73	3.71	0.0022333
rs6928703	158.25014	0.81	0.57	1.95	0.72	0.0009352
rs9458851	158.25619	0.76	0.6	1.58	1.52	0.028979
rs9347773	158.26924	0.11	0.26	5.16	2.73	0.0131155
rs7764966	158.27403	0.73	0.54	3.49	4.51	0.0153131
<u>rs724897</u>	158.27719	0.77	0.53	1.27	1.61	0.0015263
rs6916934	158.27899	0.81	0.59	1.41	1.6	0.0031945
rs6455918	158.27905	0.8	0.58	0.79	0.37	0.0029693
rs6455919	158.27946	0.82	0.59	2.42	2.09	0.0022587
rs1043863	158.2847	0.79	0.58	1.81	1.28	0.0053123

Tab. 11 - Frequenze alleliche, coefficiente di variazione e significatività dei polimorfismi presenti nel locus 6q25.1-q26 ottenuti mediante la metodica dell'array.

Dati presenti in letteratura indicano che il maggior locus di suscettibilità per il tumore del polmone sia nel cromosoma 6, più precisamente nella regione 6q23-25 (*Bailey-Wilson, 2004; You, 2009*) mentre un altro studio descrive un locus di suscettibilità nella regione 6p21 (*Wang et al, 2008*). Inoltre uno studio recente ha confermato, mediante analisi di linkage su famiglie, che la regione cromosomica 6q sarebbe coinvolta nel rischio di

sviluppare tumore polmonare. Tale rischio sarebbe inoltre maggiore in individui non fumatori, o deboli fumatori, avvalorando pertanto l'ipotesi che i fattori genetici abbiano un ruolo di rilievo nella suscettibilità al cancro del polmone. Rimane tuttavia da chiarire se e come i geni presenti in questa regione siano correlati allo sviluppo del tumore (*Amos et al., 2010*).

### ***Analisi del polimorfismo rs724897***

All'interno del locus 6q25.1-q26 abbiamo deciso di eseguire ulteriori analisi sul polimorfismo rs724897 che nell'analisi all'array ha mostrato una differenza di frequenza allelica tra il gruppo dei casi e quello dei controlli particolarmente importante ( $p=0.0015263$ ), confermato anche dalla valutazione delle frequenze alleliche reali nei DNA facenti parte dei pool. Il polimorfismo rs724897 è del tipo G/T ed è localizzato in uno degli introni del gene SNX9.

Dai risultati precedentemente descritti abbiamo osservato una differenza di frequenza allelica statisticamente significativa tra i casi di adenocarcinoma polmonare e i controlli. Per questo polimorfismo abbiamo infatti osservato un valore di odds ratio di 0.24 (CI 95% 0.08-0.74 e  $p=0.0095$ ). Così come avvenuto per il polimorfismo rs30616 precedentemente analizzato, anche per il polimorfismo rs724897 abbiamo ipotizzato che esso potesse essere un marcatore di suscettibilità allo sviluppo del tumore polmonare. Abbiamo quindi deciso di ampliare la casistica analizzandolo su ulteriori 53 casi di adenocarcinoma polmonare e su 275 individui sani come controllo. Inoltre, così come già fatto precedentemente, abbiamo deciso di analizzare il genotipo degli stessi polimorfismi su 35 carcinomi squamocellulari per ricercare l'esistenza di eventuali differenze di genotipo all'interno delle diverse tipologie di tumore polmonare. I dati relativi ai nuovi campioni esaminati sono mostrati in tabella 7.

Dai risultati ottenuti mediante analisi del genotipo al Pyrosequencing abbiamo calcolato le frequenze alleliche del polimorfismo nei gruppi studiati.

<i>Campione</i>	<i>Genotipi</i>			<i>Frequenze alleliche</i>
	<i>T/T</i>	<i>T/G</i>	<i>G/G</i>	
<i>ADCA</i>	75	17	0	0.9
<i>Carcinomi squamocellulari</i>	29	4	2	0.89
<i>Controlli</i>	217	85	12	0,77

Tab. 12 - Genotipi e frequenze alleliche dei campioni dell'intera casistica del polimorfismo rs724897.

Come è possibile osservare dalla tabella 12 non è stato osservato nessun omozigote G/G tra i casi di adenocarcinoma analizzati. Su tali dati abbiamo poi calcolato i valori di Odds Ratio per valutare l'associazione tra i genotipi dei polimorfismi analizzati e il rischio di sviluppare tumore polmonare.

<i>Modello</i>	<i>ADCA vs Controlli</i>		<i>Carcinomi squamo cellulari vs Controlli</i>	
	<i>Significatività valore-p</i>	<i>OR (CI 95%)</i>	<i>Significatività valore-p</i>	<i>OR (CI 95%)</i>
<i>TT vs TG o GG</i>	0.0198	0.51 (0.28-0.90)	0.907	0.46 (0.19-1.15)
<i>TT vs TG</i>	0.0638	0.58 (0.32-1.04)	0.0478	0.35 (0.12-1.03)
<i>TT vs GG</i>	-	-	0.7791	1.25 (0.27-5.85)

Tab.13- Significatività espressa tramite valore-p, valori di Odds Ratio e relativi Intervalli di Confidenza al 95% dei campioni dell'intera casistica per il polimorfismo rs724897.

Dall'analisi effettuata è risultato che i portatori della variante allelica rara G (omozigoti G/G e eterozigoti G/T) hanno un ridotto rischio di sviluppare adenocarcinoma polmonare (OR uguale a 0.51 con CI 95% compreso tra 0.28 e 0.90) rispetto agli omozigoti T/T. I risultati ottenuti indicano, inoltre, che anche il polimorfismo rs724897 non sia associato allo sviluppo del carcinoma squamocellulare, ha mostrato infatti un valore-p uguale a 0.907 con un valore di OR uguale 0.46 e CI 95% 0.19-1.15. In base ai risultati ottenuti possiamo quindi affermare che entrambi i polimorfismi analizzati rs30616 e rs724897 siano associati con il rischio di sviluppare l'adenocarcinoma polmonare ma non siano invece associati con lo sviluppo di carcinomi squamocellulari.

### ***Il gene SNX9***

Il polimorfismo rs724897 si trova nel gene SNX9 (sorting nexin 9), localizzato nella regione cromosomica 6q25.1-q26. Tale gene codifica per una proteina coinvolta nella regolazione dell'endocitosi cellulare. La proteina Snx9 appartiene ad una superfamiglia di proteine che hanno un ruolo diretto nello smistamento di componenti cellulari tra vari compartimenti, comprendenti i recettori transmembrana, gli endosomi e le vescicole. Le proteine appartenenti a questa famiglia hanno un dominio PX destinato al legame con il fosfatidilinositolo e un dominio BAR destinato invece alla formazione dei tubuli delle vescicole. Tuttavia due membri di tale gruppo, SNX9 e SNX18, hanno anche un dominio SH<sub>3</sub> che potrebbe quindi dirigere interazioni specifiche con un gran numero di proteine che presentano delle sequenze ricche in prolina. La proteina Snx9 è in grado di legarsi mediante il sito SH<sub>3</sub> alla proteina a funzione GTP-asi dynamine-2. Il complesso così formato si associa alle vescicole di endocitosi attraverso il residuo N-terminale del dominio SH<sub>3</sub> che è in grado di interagire anche con altri elementi, come i recettori transmembrana. Inoltre anche un'altra proteina, AP-2, si lega al complesso attraverso il motivo LC4, situato

tra i domini SH<sub>3</sub> e PX. AP-2 raccoglie il carico destinato all'endocitosi dalla membrana e lo guida fino alle vescicole. Il meccanismo di endocitosi è controllata dall'enzima aldolase che previene l'associazione del complesso SNX9:dynamine-2:AP-2 con la membrana cellulare (*Rangarajan et al., 2010*).

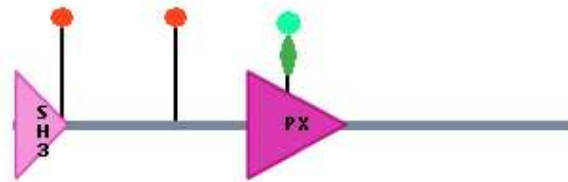


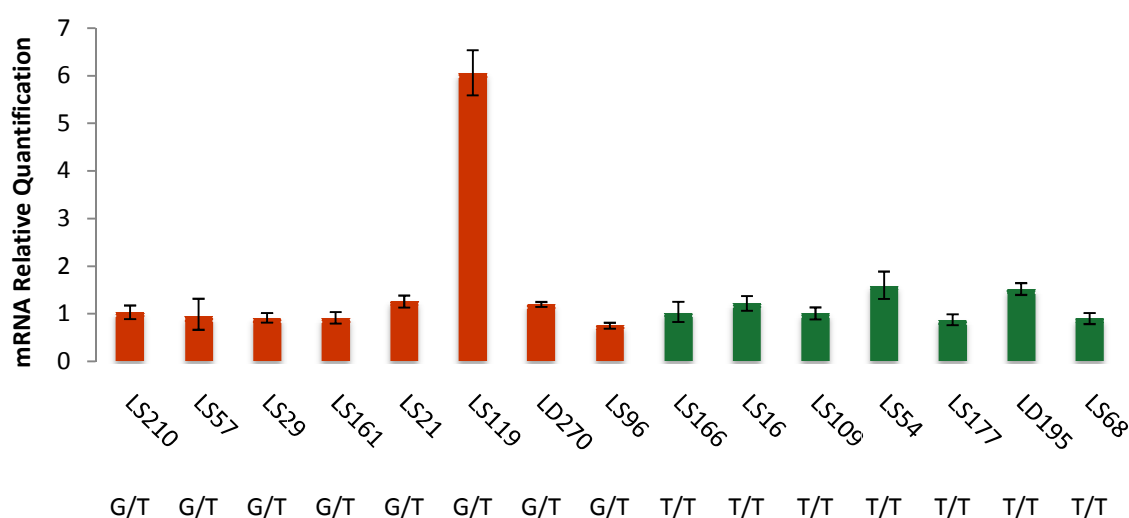
Figura 5 - Struttura della proteina SNX9

SNX9 inoltre si lega tramite il dominio SH<sub>3</sub> alla proteina ad attività tirosin-chinasica ACK2 (Activated Cdc42-associated kinase-2) che partecipa anch'essa ai meccanismi di endocitosi. ACK2 determinerebbe, mediante fosforilazione, l'oligomerizzazione di SNX9 con altre proteine della stessa famiglia. Il complesso SNX9:ACK2 sarebbe in grado di interagire con il recettore per il fattore di crescita epidermico (EGFR), riducendone i livelli nella cellula, suggerendo quindi per SNX9 un ruolo nella degradazione dell'EGFR durante l'endocitosi. Inoltre è stato osservato che il fattore di crescita epidermico (EGF) stimola la fosforilazione e quindi l'attivazione di ACK2 (*Lin et al., 2002*).

Sebbene per tale gene non esistano studi che descrivano alcuna associazione con il cancro, è noto che il recettore il fattore di crescita epidermico (EGFR) risulta overespresso nel 40-80 % dei tumori polmonari NSCLC (*Hirsch et al., 2009*), quindi alla luce di queste osservazioni non è possibile escludere un qualche tipo di coinvolgimento diretto e/o indiretto per tale gene nello sviluppo di tumore polmonare.

### *Analisi dell'espressione del gene Snx9*

Così come fatto precedentemente per il gene TRIO anche per il gene SNX9 abbiamo voluto valutare se le due varianti alleliche del polimorfismo fossero correlate a eventuali differenze di espressione genica e abbiamo quindi analizzato l'espressione del gene SNX9 in 15 campioni di tessuto normale (indicati in tabella 10), di cui 8 con il genotipo G/T e 7 con la variante polimorfica T/T.



normale con differente

Attraverso il test t abbiamo valutato se esistessero differenze di espressione genica tra i campioni con il genotipo G/T e quelli con genotipo T/T, ottenendo un valore di  $p=0.06$ , quindi non statisticamente significativo.

Per entrambi i geni analizzati, TRIO e SNX9, non abbiamo quindi evidenziato alcuna alterazione nell'espressione che ci possa permettere di correlare il genotipo dei polimorfismi individuato come associato con lo sviluppo del tumore polmonare.

Non è possibile comunque escludere la possibilità che questi stessi polimorfismi (o altri polimorfismi vicini e in linkage disequilibrium) possano esercitare un qualche effetto

modulatorio sull'espressione di questi geni. Infatti, la sonda da noi utilizzata per l'analisi di espressione genica riconosce diverse isoforme e sequenze correlate ai geni da noi analizzati.



### ***Ricerca delle isoforme dei geni TRIO e SNX9***

Abbiamo cercato nei maggiori database scientifici e in letteratura la descrizione di eventuali isoforme del gene TRIO. Nel database dell'Ensembl, per l'uomo vengono indicati 23 trascritti generati per splicing alternativo, dei quali 3 codificanti per una proteina ed una per un mRNA a decadimento non-senso e gli altri non generanti una proteina. Tali isoforme vengono ipotizzate usando un modello di previsione del database Ensembl basato sia su constatazioni sperimentali accertate, basate su diversi metodi di sequenziamento, che su dati sperimentali più flessibili, tenendo conto anche di quanto pubblicato su altri database (<http://www.ensembl.org>).

Nel secondo database consultato, quello del National Center for Biotechnology Information (NCBI), viene descritto un unico trascritto ufficiale indicato dalla sigla NM\_007118.2, tuttavia vengono poi indicate 14 sequenze correlate, delle quali 8 vengono trascritti in proteine (<http://www.ncbi.nlm.nih.gov/>).

### ***Le isoforme del gene Trio***

Nel 2006 Portales *et al.* hanno identificato nel topo 4 differenti isoforme di Trio generate mediante splicing alternativo, che si differenziano per la presenza di uno o due domini Rho-GEF (GEFDs) (Portales-Casamar *et al.*, 2006.)

In cellule-T adulte di leucemia (ATL) di alcuni pazienti è stato identificato un gene chiamato *Tgat* (*Trio*-related transforming gene in ATL tumors cells) generato dallo splicing alternativo tra la sequenza del gene Trio che codifica per il dominio Rho-GEF (dall'esone 38 al 46) e un nuovo esone, (esone 58), localizzato a valle dell'ultimo esone di Trio codificante per una sequenza di 15 aminoacidi (Yoshizuka N. *et al.*, 2004).

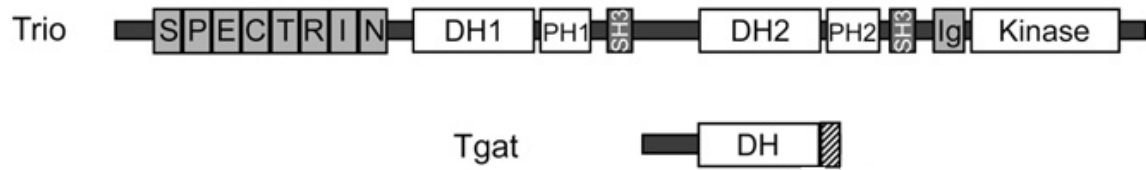


Fig.8- Rappresentazione schematica di Trio e della sua isoforma Tgat

L'espressione di Tgat è stata descritta nelle cellule ATL, mentre non è stata rilevata nei linfociti normali, suggerendone un ruolo nel fenotipo maligno in questo tipo di leucemia. La sua espressione nella linea cellulare di fibroblasti murini ha inoltre messo in luce la natura di oncoproteina di Tgat esercitata attraverso l'inibizione delle metalloproteinasi MMP (*Mori T. et al., 2007*). Yamada e collaboratori hanno dimostrato che in cellule T adulte di leucemia (ATL) Tgat attiva il fattore di trascrizione NF-kB, attraverso il residuo C-terminale di cui solo tale isoforma è dotata (*Yamada Y. et al., 2007*).

### ***Le isoforme del gene SNX9***

Così come fatto precedentemente per il gene TRIO, anche per il gene SNX9 abbiamo cercato l'esistenza di eventuali isoforme prima nei maggiori database informatici e poi ci siamo accertati se tali isoforme fossero state descritte in letteratura, non trovando però pubblicazioni inerenti tale argomento. Nel database dell'Ensembl invece vengono indicate due differenti isoforme del gene SNX9, entrambe tradotte in proteine (<http://www.ensembl.org>).

Nel sito del NCBI invece viene indicato un unico trascritto (NM\_016224.3) ma vengono però poi indicate numerose sequenze correlate a quella unicamente descritta per il gene SNX9.

L'esistenza di tali isoforme ci suggerisce quindi di effettuare nuove analisi rivolte alla ricerca delle altre forme di espressione dei geni TRIO e SNX9 attraverso il disegno di primer specifici in grado di amplificare le sequenze geniche di interesse e discriminare tra la varie isoforme esistenti.

## ***Conclusioni***

Alla luce di tali risultati possiamo affermare che la ricerca delle varianti genetiche disperse sull'intero genoma associate allo sviluppo di patologie complesse possono essere condotte su pool costituiti da diversi campioni di DNA appartenenti a differenti individui, senza che questo possa incidere sull'efficacia dell'esperimento. Questo tipo di approccio viene quindi confermato come particolarmente utile negli studi di associazione caso-controllo volti alla ricerca delle varianti genetiche disperse sull'intero genoma, così come già alcuni lavori in letteratura avevano indicato (*Pearson, 2007; Zhang, 2008*).

Abbiamo individuato diversi loci che presentano numerosi polimorfismi con differenze di frequenze alleliche statisticamente significative tra il gruppo dei casi e quello dei controlli. Tale risultato conferma quanto descritto in letteratura, cioè che la suscettibilità allo sviluppo del tumore polmonare sia un'ereditarietà di tipo poligenico, con il coinvolgimento di più loci genetici (*Peto J. and Houlston RS., 2001; Dragani TA, 2003*). Tra i loci individuati due in particolare hanno mostrato un'elevata densità di polimorfismi con differenze di frequenze alleliche statisticamente significative tra il gruppo dei casi e quello dei controlli localizzati nelle posizioni 5p15.2 e 6q25.1-q26. Tali regioni sono adiacenti ad alcuni loci di suscettibilità al tumore polmonare già descritti in letteratura nelle posizioni 5p15.33 (*Wang et al, 2008; McKay et al., 2008*) e 6p21 (*Wang et al, 2008*). Il maggior locus di suscettibilità viene invece indicato nella regione 6q23-25 (*Bailey-Wilson et al., 2004; You et al., 2009*). All'interno di questi due loci abbiamo individuato due polimorfismi, rs30616 (localizzato sul cromosoma 5 all'interno del gene TRIO) e rs724897 (sul cromosoma 6 all'interno del gene SNX9), che presentano differenza di frequenza allelica statisticamente significativa tra le più alte tra tutti i polimorfismi all'interno dei loci analizzati. Dalle analisi sui genotipi che abbiamo eseguito sui polimorfismi rs30616 e rs724897 possiamo affermare che per entrambi la variante allelica rara sia associata con un ridotto rischio di sviluppare adenocarcinoma polmonare. Non abbiamo ottenuto però il

medesimo risultato con i carcinomi squamocellulari per i quali non abbiamo riscontrato correlazione con il genotipo dei polimorfismi rs30616 e rs724897. I risultati riportati mostrano che entrambi i polimorfismi analizzati risultano essere associati al rischio di sviluppare l'adenocarcinoma polmonare quindi entrambi i loci in esame, o loci ad essi vicini in linkage disequilibrium, correlano con la suscettibilità alla malattia. Ulteriori studi con un maggior numero di campioni, saranno necessari per supportare l'ipotesi che questi polimorfismi siano associati con lo sviluppo dell'adenocarcinoma polmonare.

Le analisi di espressione dei geni TRIO e SNX9 non evidenziano alcuna alterazione che ci possa permettere di correlare l'allele dei polimorfismi individuato come associato allo sviluppo dell'adenocarcinoma polmonare con il fenotipo tumorale. Inoltre non abbiamo riscontrato differenze di espressione per entrambi i geni analizzati, TRIO e SNX9. Ad oggi non sono riportate alterazioni biologiche (espressione di isoforme diverse, disregolazioni epigenetiche o alterazioni nella sequenza dei geni) a carico dei geni TRIO ed SNX9 correlabili agli alleli dei polimorfismi analizzati. Le ipotesi che ne derivano comprendono sia la possibilità che tali alterazioni debbano ancora essere individuate sia che questi polimorfismi non siano direttamente correlati al rischio di sviluppare l'adenocarcinoma polmonare ma potrebbero essere in linkage disequilibrium con altri marcatori direttamente associati allo sviluppo di tale patologia.

## ***Bibliografia***

**Abecasis** GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO. *Extent and distribution of linkage disequilibrium in three genomic regions*. Am J Hum Genet. 2001 Jan;68(1):191-197. Epub 2000 Nov 13.

**Ahlbom** A, Lichtenstein P, Malmstrom H, Feychting M, Hemminki K, Pedersen NL. *Cancer in twins: genetic and nongenetic familial risk factors*. J Natl Cancer Inst., 89:287-93, 1997

**Ahrendt** SA, Decker PA, Alawi EA, Zhu Yr YR, Sanchez-Cespedes M, Yang SC, Haasler GB, Kajdacsy-Balla A, Demeure MJ, Sidransky D. *Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung*. Cancer. 2001 Sep 15;92(6):1525-30.

**AIRTUM** Working Group, Crocetti E, Buzzoni C. *New incidence and mortality data. 2003-2005*. Epidemiol Prev. 2009 Apr 23;33(1-2 Suppl 2):e1-3, e5-26. Italian.

**Ameur** A, Rada-Iglesias A, Komorowski J, Wadelius C. *Identification of candidate regulatory SNPs by combination of transcription-factor-binding site prediction, SNP genotyping and haploChIP*. Nucleic Acids Res. 2009 Jul;37(12):e85. Epub 2009 May 18.

**Amos** CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. *Genome-wide association scan of tag SNPs*



*identifies a susceptibility locus for lung cancer at 15q25.1.* Nat Genet. 2008 May;40(5):616-22. Epub 2008 Apr 2.

**Amos** CI, Pinney SM, Li Y, Kupert E, Lee J, de Andrade MA, Yang P, Schwartz AG, Fain PR, Gazdar A, Minna J, Wiest JS, Zeng D, Rothschild H, Mandal D, You M, Coons T, Gaba C, Bailey-Wilson JE, Anderson MW. *A susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers.* Cancer Res. 2010 Mar 15;70(6):2359-67. Epub 2010 Mar 9.

**Anderson** PD, Huizing M, Claassen DA, White J, Gahl WA. *Hermansky-Pudlak syndrome type 4 (HPS-4): clinical and molecular characteristics.* Hum Genet. 2003 Jul;113(1):10-7. Epub 2003 Mar 27.

**Bailey-Wilson** JE, Amos CI, Pinney SM, Petersen GM, de Andrade M, Wiest JS, Fain P, Schwartz AG, You M, Franklin W, Klein C, Gazdar A, Rothschild H, MandalD, Coons T, Slusser J, Lee J, Gaba C, Kupert E, Perez A, Zhou X, Zeng D, Liu Q, Zhang Q, Seminara D, Minna J, Anderson MW. *A major lung cancer susceptibility locus maps to chromosome 6q23-25.* Am J Hum Genet. 2004 Sep;75(3):460-74. Epub 2004 Jul 21.

**Balsara** BR, Sonoda G, du Manoir S, Siegfried JM, Gabrielson E, Testa JR. *Comparative genomic hybridization analysis detects frequent, often high-level, overrepresentation of DNA sequences at 3q, 5p, 7p, and 8q in human non-small cell lung carcinomas.* Cancer Res. 1997 Jun 1;57(11):2116-20.

**Barratt** BJ, Payne F, Rance HE, Nutland S, Todd JA, Clayton DG. *Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design.* Ann Hum Genet. 2002 Nov;66(Pt 5-6):393-405.

**Bateman** J, Van Vactor D. *The Trio family of guanine-nucleotide-exchange factors: regulators of axon guidance.* J Cell Sci. 2001 Jun;114(Pt 11):1973-80.

**Baud** V, Karin M. *Is NF-kappaB a good target for cancer therapy? Hopes and pitfalls.* Nat Rev Drug Discov. 2009 Jan;8(1):33-40. Review. PubMed PMID: 19116625; PubMed Central PMCID: PMC2729321

**Bernig** T, Chanock SJ. *Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer.* Expert Rev Mol Diagn. 2006 May;6(3):319-31.

**Bhattacharjee** A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, Meyerson M. *Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses.* Proc Natl Acad Sci U S A. 2001 Nov 20;98(24):13790-5. Epub 2001 Nov 13.

**Bishop** JM. *Oncogenes and proto-oncogenes.* J Cell Physiol Suppl. 1986;4:1-5.

**Bignell** GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, Buck G, Chen L, Beare D, Latimer C, Widaa S, Hinton J, Fahey C, Fu B, Swamy S, Dalgliesh GL, Teh BT, Deloukas P, Yang F, Campbell PJ, Futreal PA, Stratton MR. *Signatures of mutation and selection in the cancer genome.* Nature. 2010 Feb 18;463(7283):893-8.

**Brandon M, Baldi P, Wallace DC.** *Mitochondrial mutations in cancer.* Oncogene 2006 Aug 7;25(34):4647-62.

**Brookes AJ.** *The essence of SNPs.* Gene. 1999 Jul 8;234(2):177-86.

**Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA.** *Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans.* Nat Genet. 2003 Apr;33(4):518-21. Epub 2003 Mar 24

**Carlson CS, Eberle MA, Kruglyak L, Nickerson DA.** *Mapping complex disease loci in whole-genome association studies.* Nature. 2004 May 27;429(6990):446-52.

**Chanoock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS.** *Replicating genotype-phenotype associations.* Nature. 2007 Jun

**Chattopadhyay I, Singh A, Phukan R, Purkayastha J, Kataki A, Mahanta J, Saxena S, Kapur S.** *Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma exposed to tobacco and betel quid from high-risk area in India.* Mutat Res. 2010 Feb 2;696(2):130-138. Epub 2010 Jan 18.

**Cheng JC**, Tang NL, Yeung HY, Miller N. *Genetic association of complex traits: using idiopathic scoliosis as an example*. Clin Orthop Relat Res. 2007 Sep;462:38-44.

**Coe BP**, Henderson LJ, Garnis C, Tsao MS, Gazdar AF, Minna J, Lam S, Macaulay C, Lam WL. *High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines*. Genes Chromosomes Cancer. 2005 Mar;42(3):308-13.

**Dawson E**, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibbling T, Tinsley E, KirbyS, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Easton DF, Pooley KA, Dunning AM, et al. *Genome-wide association study identifies novel breast cancer susceptibility loci*. Nature. 2007;447(7148):1087-1093.

**Dragani TA**, Canzian F, Pierotti MA. *A polygenic model of inherited predisposition to cancer*. FASEB J. 1996 Jun;10(8):865-70.

**Dragani TA**. *10 years of mouse cancer modifier loci: human relevance*. Cancer Res. 2003 Jun 15;63(12):3011-8.

**Dubey S**, Powell CA. *Update in lung cancer 2006*. Am J Respir Crit Care Med. 2007 May 1;175(9):868-74.

**Elston RC**. *The genetic dissection of multifactorial traits*. Clin Exp Allergy. 1995 Nov;25 Suppl 2:103-6.

**Fearon ER**. *Human cancer syndromes: clues to the origin and nature of cancer*. Science. 1997 Nov 7;278(5340):1043-50.

**Fernandez E**, La Vecchia C, D'Avanzo B, Negri E, Franceschi S. *Family history and the risk of liver, gallbladder, and pancreatic cancer*. Cancer Epidemiol Biomarkers Prev. 1994 Apr-May;3(3):209-12.

**Frayling TM**, McCarthy MI. *Genetic studies of diabetes following the advent of the genome-wide association study: where do we go from here?* Diabetologia. 2007 Nov;50(11):2229-33 .

**Freedman ND**, Leitzmann MF, Hollenbeck AR, Schatzkin A, Abnet CC. *Cigarette smoking and subsequent risk of lung cancer in men and women: analysis of a prospective cohort study*. Lancet Oncol. 2008 Jul;9(7):649-56. Epub 2008 Jun 13.

**Ford D**, Easton DF, Peto J. *Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence*. Am J Hum Genet. 1995 Dec;57(6):1457-62.

**Foulds L**. *The experimental study of tumor progression: a review*. Cancer Res.1954 Jun;14(5):327-39.

**Gabriel SB**, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, LochnerA, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ,Altshuler D. *The structure of haplotype blocks in the human genome*. Science 2002; 296: 2225-9.

**Galvan** A, Ioannidis JP, Dragani TA. *Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer*. Trends Genet. 2010 Mar;26(3):132-41. Epub 2010 Jan 26.

**Garber** ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D, Petersen I. *Diversity of gene expression in adenocarcinoma of the lung*. Proc Natl Acad Sci U S A. 2001 Nov 20;98(24):13784-9. Epub 2001 Nov 13. Erratum in: Proc Natl Acad Sci U S A 2002 Jan 22;99(2):1098.

**Garnis** C, Davies JJ, Buys TP, Tsao MS, MacAulay C, Lam S, Lam WL. *Chromosome 5p aberrations are early events in lung cancer: implication of glial cell line-derived neurotrophic factor in disease progression*. Oncogene. 2005 Jul 14;24(30):4806-12.

**Goldstein**, Ahmadi, Weale, Wood, *Genome scans and candidate gene approaches in the study of common diseases and variable drug responses*. Trends Genet., 2003; 19: 615–622

**Govindan** R., Page N, Morgensztern D, Read W, Tierney R, Vlahiotis A, Spitznagel EL, Piccirillo J. *Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database*. J Clin Oncol. 2006 Oct 1;24(28):4539-44.

**Gower** S., Hammond D. *CSP deposition to the alveolar region of the lung: implications of cigarette design*. Risk Anal. 2007 Dec;27(6):1519-33.

**Greenman** C., Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR. *Patterns of somatic mutation in human cancer genomes*. Nature. 2007 Mar 8;446(7132):153-8.

**Gunderson** KL, Kuhn KM, Steemers FJ, Ng P, Murray SS, Shen R. *Whole-genome genotyping of haplotype tag single nucleotide polymorphisms*. Pharmacogenomics. 2006 Jun;7(4):641-8.

**Guo** SW. *Gene-environment interaction and the mapping of complex traits: some statistical models and their implications*. Hum Hered. 2000 Sep-Oct;50(5):286-303.

**Haber** DA, Settleman J. *Cancer: drivers and passengers*. Nature. 2007 Mar 8;446(7132):145-6.

**Hanahan** D, Weinberg RA. *The hallmarks of cancer*. Cell. 2000 Jan 7;100(1):57-70.

**Harris** RC, Chung E, Coffey RJ. *EGF receptor ligands*. Exp Cell Res. 2003 Mar 10;284(1):2-13.

**Herbst** RS, Heymach JV, Lippman SM. *Lung cancer*. N Engl J Med. 2008 Sep 25;359(13):1367-80

**Hersh** CP, Pillai SG, Zhu G, Lomas DA, Bakke P, Gulsvik A, DeMeo DL, Klanderman BJ, Lazarus R, Litonjua AA, Sparrow D, Reilly JJ, Agusti A, Calverley PM, Donner CF, Levy RD, Make BJ, Paré PD, Rennard SI, Vestbo J, Wouters EF, Scholand MB, Coon H, Hoidal J, Silverman EK. *Multistudy fine mapping of chromosome 2q identifies XRCC5 as a chronic obstructive pulmonary disease susceptibility gene*. Am J Respir Crit Care Med. 2010 Sep 1;182(5):605-13. Epub 2010 May 12.

**Hibi** K, Mizukami H, Shirahata A, Goto T, Sakata M, Sanada Y. *Aberrant methylation of the netrin-1 receptor genes UNC5C and DCC detected in advanced colorectal cancer*. World J Surg. 2009 May;33(5):1053-7.

**Hirsch** FR, Varella-Garcia M, Cappuzzo F. *Predictive value of EGFR and HER2 overexpression in advanced non-small-cell lung cancer*. Oncogene. 2009 Aug;28 Suppl 1:S32-7.

**Hoffmann** D, Rivenson A, Hecht SS. *The biological significance of tobacco-specific N-nitrosamines: smoking and adenocarcinoma of the lung*. Crit Rev Toxicol. 1996;26(2):199-211.



**Hoover RN.** *The evolution of epidemiologic research: from cottage industry to "big" science.* Epidemiology. 2007 Jan;18(1):13-7.

**Horazdovsky BF, Davies BA, Seaman MN, McLaughlin SA, Yoon S, Emr SD.** *A sorting nexin-1 homologue, Vps5p, forms a complex with Vps17p and is required for recycling the vacuolar protein-sorting receptor.* Mol Biol Cell. 1997 Aug;8(8):1529-41.

**Hunter DJ, Kraft P.** *Drinking from the fire hose statistical issues in genome-wide association studies.* N Engl J Med. 2007;357(5):436-439

**Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A, McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, Clavel-Chapelon F, Vineis P, Bueno-de-Mesquita HB, Lund E, Martinez C, Bingham S, Rasmuson T, Hainaut P, Riboli E, Ahrens W, Benhamou S, Lagiou P, Trichopoulos D, Holcátová I, Merletti F, Kjaerheim K, Agudo A, Macfarlane G, Talamini R, Simonato L, Lowry R, Conway DI, Znaor A, Healy C, Zelenika D, Bolland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P.** *A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25.* Nature. 2008 Apr 3;452(7187):633-7.

**International HapMap Consortium.** *A second generation human haplotype map of over 3.1 million SNPs.* Nature. 2007 Oct 18;449(7164):851-61.

**Jemal** A, Siegel R, Xu J, Ward E. *Cancer statistics, 2010*. CA Cancer J Clin. 2010 Sep-Oct;60(5):277-300. Epub 2010 Jul 7.

**Jemal** A., Siegel R, Ward E, Murray T, Xu J, Smigal C, Thun MJ. *Cancer statistics, 2006*. CA Cancer J Clin. 2006 Mar-Apr;56(2):106-30.

.

**Jemal** A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. *Cancer statistics, 2007*. CA Cancer J Clin. 2007 Jan-Feb;57(1):43-66.

**Jiang** WG, Watkins G, Lane J, Cunnick GH, Douglas-Jones A, Mokbel K, Mansel RE. *Prognostic value of rho GTPases and rho guanine nucleotide dissociation inhibitors in human breast cancers*. Clin Cancer Res. 2003 Dec 15;9(17):6432-40.

**Jin** H, Evangelou E, Ioannidis JP, Ralston SH. *Polymorphisms in the 5' flank of COL1A1 gene and osteoporosis: meta-analysis of published studies*. Osteoporos Int. 2010 Aug 27.

**Johns** LE, Houlston RS. *A systematic review and meta-analysis of familial colorectal cancer risk*. Am J Gastroenterol. 2001 Oct;96(10):2992-3003.

**Johnson** GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. *Haplotype tagging for the identification of common disease genes*. Nat Genet. 2001 Oct;29(2):233-7.

**Karin M.** *Nuclear factor-kappaB in cancer development and progression.* Nature. 2006 May 25;441(7092):431-6. Review. PubMed PMID: 16724054.

**Kinzler KW, Vogelstein B.** *Cancer-susceptibility genes. Gatekeepers and caretakers.* Nature. 1997 Apr 24;386(6627):761, 763.

**Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC.** *A genome-wide association study in 574 schizophrenia trios using DNA pooling.* Mol Psychiatry. 2009 Aug;14(8):796-803.Epub 2008 Mar 11.

**Knudson AG.** *Antioncogenes and human cancer.* Proc Natl Acad Sci U S A. 1993 Dec 1;90(23):10914-21.

**Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T.** *Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications.* Cancer Res. 2004 Dec 15;64(24):8919-23.

**Kruglyak L., Nickerson DA.** *Variation is the spice of life.* Nat Genet. 2001 Mar;27(3):234-6

**Lander ES, Schork NJ.** *Genetic dissection of complex traits.* Science. 1994 Sep 30;265(5181):2037-48. Review. Erratum in: Science 1994 Oct 21;266(5184):353.

**Lane J.**, Martin TA, Mansel RE, Jiang WG. *The expression and prognostic value of the guanine nucleotide exchange factors (GEFs) Trio, Vav1 and TIAM-1 in human breast cancer.* Int Semin Surg Oncol. 2008 Oct 16;5:23.

**Lengauer C.**, Kinzler KW, Vogelstein B. *Genetic instabilities in human cancers.* Nature. 1998 Dec 17;396(6712):643-9.

**Libioulle C**, Louis E, Hansoul S, et al. *Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4.* PLoS Genet. 2007;3(4):e58

**Lichtenstein P**, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland.* N Engl J Med. 2000 Jul 13;343(2):78-85.

**Lin Q**, Lo CG, Cerione RA, Yang W. *The Cdc42 target ACK2 interacts with sorting nexin 9 (SH3PX1) to regulate epidermal growth factor receptor degradation.* J Biol Chem. 2002 Mar 22;277(12):10134-8.

**Loeb LA.** *Mutator phenotype may be required for multistage carcinogenesis.* Cancer Res. 1991 Jun 15;51(12):3075-9.

**Loeb** LA, Loeb KR, Anderson JP. *Multiple mutations and cancer*. Proc Natl Acad Sci U S A. 2003 Feb 4;100(3):776-81. Epub 2003 Jan 27.

**Lohmussaar** E, Zernant J, Tonisson N, RemmM, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, CardonLR, Dunham I. *A first-generation linkage disequilibrium map of human chromosome 22*. Nature, 2002;418: 544–548

**Luk** C, Tsao MS, Bayani J, Shepherd F, Squire JA. *Molecular cytogenetic analysis of non-small cell lung carcinoma by spectral karyotyping and comparative genomic hybridization*. Cancer Genet Cytogenet. 2001 Mar;125(2):87-99.

**Macgregor** S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW, Visscher PM. *Highly cost-efficient genome-wide association studies using DNA pool and dense SNP arrays*. Nucleic Acids Res. 2008 Apr;36(6):e35. Epub 2008 Feb 14.

**Manolio** TA, Bailey-Wilson JE, Collins FS. *Genes, environment and the value of prospective cohort studies*. Nat Rev Genet. 2006;7(10):812-820.

**McKay** JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpén F, Elvestad MB, Vatten L, Njølstad I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, Lubiński J, Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T,

Xinarianos G, Cassidy A; EPIC Study, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, González CA, Ramón Quirós J, Martínez C, Navarro C, Ardanaz E, Larrañaga N, Kham KT, Key T, Bueno-de-Mesquita HB, Peeters PH, Trichopoulou A, Linseisen J, Boeing H, Hallmans G, Overvad K, Tjønneland A, Kumle M, Riboli E, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P. *Lung cancer susceptibility locus at 5p15.33*. Nat Genet. 2008 Dec;40(12):1404-6. Epub 2008 Nov 2.

**Medley** QG, Buchbinder EG, Tachibana K, Ngo H, Serra-Pagès C, Streuli M. *Signaling between focal adhesion kinase and trio*. J Biol Chem. 2003 Apr 11;278(15):13265-70. Epub 2003 Jan 27.

**Meyerson** M., Franklin, W.A., and Kelley, M.J. 2004. *Molecular classification and molecular genetics of human lung cancers*. Semin. Oncol. 31:4–19

**Michor** F, Iwasa Y, Nowak MA. *Dynamics of cancer progression*. Nat Rev Cancer. 2004 Mar;4(3):197-205.

**Mori** T, Moriuchi R, Okazaki E, Yamada K, Katamine S. *Tgat oncoproteina functions as a inhibitor of RECK by association of the unique C-terminal region*. Biochem Biophys Res Commun. 2007 Apr 20;355(4):937-43. Epub 2007 Feb 20.

**Morley** M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. *Genetic analysis of genome-wide variation in human gene expression*. Nature 2004 Aug 12;430(7001):743-7. Epub 2004 Jul 21.

**Nachman** MW, Crowell SL. *Estimate of the mutation rate per nucleotide in humans. Genetics.* 2000 Sep;156(1):297-304.

**Naoki** K, Chen TH, Richards WG, Sugarbaker DJ, Meyerson M. *Missense mutations of the BRAF gene in human lung adenocarcinoma. Cancer Res.* 2002 Dec 1;62(23):7001-3.

**National Institutes of Health.** *Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS).* Federal Regist. 2007;72(166):49290-49297.

**Pääbo** S, *The mosaic that is our genome. Nature.* 2003 Jan 23;421(6921):409-12.

**Patil** N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP. *Blocks of limited haplotype diversity revealed by high resolution scanning of human chromosome 21. Science* 2001; 294: 1719–1723

**Pearson** JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, Webster JA, Beach T, Sando SB, Aasly JO, HeunR, Jessen F, Kolsch H, Tsolaki M, Daniilidou M, Reiman EM, Papassotiropoulos A, Hutton ML, Stephan DA, Craig DW. *Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. Am J Hum Genet.* 2007 Jan;80(1):126-39. Epub 2006 Dec 6.

**Pearson** TA, Manolio TA. *How to interpret a genome-wide association study*. JAMA. 2008 Mar 19;299(11):1335-44. Erratum in: JAMA. 2008 May 14;299(18):2150

**Peto** J, Houlston RS. *Genetics and the common cancers*. Eur J Cancer. 2001 Oct;37 Suppl 8:S88-96.

**Pharoah** PD, Day NE, Duffy S, Easton DF, Ponder BA. *Family history and the risk of breast cancer: a systematic review and meta-analysis*. Int J Cancer. 1997 May 29;71(5):800-9.

**Ponder** BA. *Genetic predisposition to cancer*. Br J Cancer. 1991 Aug;64(2):203-4.

**Popescu** NC, Zimonjic DB. *Molecular cytogenetic characterization of cancer cell alterations*. Cancer Genet Cytogenet. 1997 Jan;93(1):10-21.

**Portales-Casamar** E, Briançon-Marjollet A, Fromont S, Triboulet R, Debant A. *Identification of novel neuronal isoforms of the Rho-GEF Trio*. Biol Cell. 2006 Mar;98(3):183-93.

**Rangarajan** ES, Park H, Fortin E, Sygusch J, Izard T. *Mechanism of aldolase control of sorting nexin 9 function in endocytosis*. J Biol Chem. 2010 Apr 16;285(16):11983-90. Epub 2010 Feb 2.



**Reich** DE, Lander ES. *On the allelic spectrum of human disease*. Trends genet. 2001; 17: 502-10.

**Ried** T, Petersen I, Holtgreve-Grez H, Speicher MR, Schröck E, du Manoir S, Cremer T. *Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization*. Cancer Res. 1994 Apr 1;54(7):1801-6.

**Riely** GJ, Politi KA, Miller VA, Pao W. *Update on epidermal growth factor receptor mutations in non-small cell lung cancer*. Clin Cancer Res. 2006 Dec 15;12(24):7232-41.

**Risch** N, Merikangas K. *The future of genetic studies of complex human diseases*. Science. 1996 Sep 13;273(5281):1516-7.

**Risch** NJ. *Searching for genetic determinants in the new millennium*. Nature. 2000 Jun 15;405(6788):847-56.

**Risch** N. *The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches*. Cancer Epidemiol Biomarkers Prev. 2001 Jul;10(7):733-41.

**Risch** A, Plass C. *Lung cancer epigenetics and genetics*. Int J Cancer. 2008 Jul 1;123(1):1-7.

**Renan MJ.** *How many mutations are required for tumorigenesis? Implications from human cancer data.* Mol Carcinog. 1993;7(3):139-46.

**Rodenhuis S, Slebos RJ.** *Clinical significance of ras oncogene activation in human lung cancer.* Cancer Res. 1992 May 1;52(9 Suppl):2665s-2669s.

**Santarosa M, Ashworth A.** *Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way.* Biochim Biophys Acta. 2004 Jun 7;1654(2):105-22.

**Salhia B, Tran NL, Chan A, Wolf A, Nakada M, Rutka F, Ennis M, McDonough S, Berens ME, Symons M, Rutka JT.** *The guanine nucleotide exchange factors trio, Ect2, and Vav3 mediate the invasive behavior of glioblastoma.* Am J Pathol. 2008 Dec;173(6):1828-38. Epub 2008 Nov 13.

**Sebastiani P, Zhao Z, Abad-Grau MM, Riva A, Hartley SW, Sedgewick AE, Doria A, Montano M, Melista E, Terry D, Perls TT, Steinberg MH, Baldwin CT.** *A hierarchical and modular approach to the discovery of robust associations in genome-wide association studies from pooled DNA samples.* BMC Genet. 2008 Jan 14;9:6.

**Skol AD, Scott LJ, Abecasis GR, Boehnke M.** *Optimal designs for two-stage genome-wide association studies.* Genet Epidemiol. 2007 Nov;31(7):776-88.

**Sellers**, T.A., Potter, J.D., Bailey-Wilson, J.E., Rich, S.S., Rothschild, H., and Eltson, R.C. Lung cancer detection and prevention: *Evidence for an interaction between smoking and genetic predisposition*. Cancer Res., 52: 2694s-2697s, 1992.

**Sharma S**, Kelly TK, Jones PA. *Epigenetics in cancer Carcinogenesis*. 2010 Jan;31(1):27-36. Epub 2009 Sep 13.

**Shigematsu** H, Gazdar AF. *Somatic mutations of epidermal growth factor receptorsignaling pathway in lung cancers*. Int J Cancer. 2006 Jan 15;118(2):257-62.

**Slebos** RJ, Kibbelaar RE, Dalesio O, Kooistra A, Stam J, Meijer CJ, Wagenaar SS, Vanderschueren RG, van Zandwijk N, Mooi WJ, et al. *K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung*. N Engl J Med. 1990 Aug 30;323(9):561-5.

**Stahel** RA. *Adenocarcinoma, a molecular perspective*. Ann Oncol. 2007 Jul;18 Suppl 9:ix147-9

**Steer** S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, Harrison AA, Highton J, Jones PB, O'Donnell J, Stamp L, Fitzgerald L, Iliev D, Kouzmine A, Tran T, Skolnick MH, Timms KM, Lanchbury JS, Merriman TR. *Genomic DNA pooling for whole-genome association scans in complex*

*disease: empirical demonstration of efficacy in rheumatoid arthritis.* Genes Immun. 2007 Jan;8(1):57-68. Epub 2006 Dec 7.

**Stratton** MR, Campbell PJ, Futreal PA. *The cancer genome.* Nature. 2009 Apr 9;458(7239):719-24.

**Sugio** K, Ishida T, Yokoyama H, Inoue T, Sugimachi K, Sasazuki T. *ras gene mutations as a prognostic marker in adenocarcinoma of the human lung without lymph node metastasis.* Cancer Res. 1992 May 15;52(10):2903-6.

**Sun** S, Schiller JH, Spinola M, Minna JD. *New molecularly targeted therapies for lung cancer.* J Clin Invest. 2007 Oct;117(10):2740-50.

**Suzuki** C, Takahashi K, Hayama S, Ishikawa N, Kato T, Ito T, Tsuchiya E, Nakamura Y, Daigo Y. *Identification of Myc-associated protein with JmjC domain as a novel therapeutic target oncogene for lung cancer.* Mol Cancer Ther. 2007 Feb;6(2):542-51.

**Tabor** HK, Risch NJ, Myers RM. *Candidate-gene approaches for studying complex genetic traits: practical considerations.* Nat Rev Genet. 2002 May;3(5):391-7.

**Talbot** SJ, Crawford DH. *Viruses and tumours--an update.* Eur J Cancer. 2004 Sep;40(13):1998-2005.

**Thomas** DC, Witte JS. *Point: population stratification: a problem for case-control studies of candidate-gene associations?* Cancer Epidemiol Biomarkers Prev. 2002 Jun;11(6):505-12.

**Ullmann** R, Schwendel A, Klemen H, Wolf G, Petersen I, Popper HH. *Unbalanced chromosomal aberrations in neuroendocrine lung tumors as detected by comparative genomic hybridization.* Hum Pathol. 1998 Oct;29(10):1145-9.

**Umar** A, Boyer JC, Thomas DC, Nguyen DC, Risinger JI, Boyd J, Ionov Y, Perucho M, Kunkel TA. *Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability.* J Biol Chem. 1994 May 20;269(20):14367-70.

**Vogelstein** B, Kinzler KW. *Cancer genes and the pathways they control.* Nat Med. 2004 Aug;10(8):789-99.

**Wacholder** S, Rothman N, Caporaso N. *Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer.* Cancer Epidemiol Biomarkers Prev. 2002 Jun;11(6):513-20.

**Wang** Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI, Houlston RS. *Common 5p15.33 and 6p21.33 variants influence lung cancer risk.* Nat Genet. 2008 Dec;40(12):1407-9. Epub 2008 Nov 2.

**Wei** EK, Wolin KY, Colditz GA. *Time course of risk factors in cancer etiology and progression*. J Clin Oncol. 2010 Sep 10;28(26):4052-7. Epub 2010 Jul 19.

**Weinberg** RA. *How cancer arises*. Sci Am. 1996 Sep;275(3):62-70.

**Weir** B, Zhao X, Meyerson M. *Somatic alterations in the human cancer genome*. Cancer Cell. 2004 Nov;6(5):433-8.

**Wood** ME, Kelly K, Mullineaux LG, Bunn PA Jr. *The inherited nature of lung cancer: a pilot study*. Lung Cancer, 30:135-44, 2000.

**Yamada** K, Moriuchi R, Mori T, Okazaki E, Kohno T, Nagayasu T, Matsuyama T, Katamine S. *Tgat, a Rho-specific guanine nucleotide exchange factor, activates NF-kappaB via physical association with IkappaB kinase complexes*. Biochem Biophys Res Commun. 2007 Mar 30;355(1):269-74. Epub 2007 Feb 2. PubMed PMID: 17292329.

**Yoshizuka** N, Moriuchi R, Mori T, Yamada K, Hasegawa S, Maeda T, Shimada T, Yamada Y, Kamihira S, Tomonaga M, Katamine S. *An alternative transcript derived from the trio locus encodes a guanosine nucleotide exchange factor with mouse cell-transforming potential*. J Biol Chem. 2004 Oct 15;279(42):43998-4004. Epub 2004 Aug 11. PubMed PMID: 15308664.

**You** M, Wang D, Liu P, Vikis H, James M, Lu Y, Wang Y, Wang M, Chen Q, Jia D, Liu Y, Wen W, Yang P, Sun Z, Pinney SM, Zheng W, Shu XO, Long J, Gao YT, Xiang YB, Chow WH, Rothman N, Petersen GM, de Andrade M, Wu Y, Cunningham JM, Wiest

JS, Fain PR, Schwartz AG, Girard L, Gazdar A, Gaba C, Rothschild H, Mandal D, Coons T, Lee J, Kupert E, Seminara D, Minna J, Bailey-Wilson JE, Amos CI, Anderson MW. *Fine mapping of chromosome 6q23-25 region in familial lung cancer families reveals RGS17 as a likely candidate gene.* Clin Cancer Res. 2009 Apr 15;15(8):2666-74. Epub 2009 Apr 7.

**Yoshizuka** N, Moriuchi R, Mori T, Yamada K, Hasegawa S, Maeda T, Shimada T, Yamada Y, Kamihira S, Tomonaga M, Katamine S. *An alternative transcript derived from the trio locus encodes a guanosine nucleotide exchange factor with mouse cell-transforming potential.* J Biol Chem. 2004 Oct 15;279(42):43998-4004. Epub 2004 Aug 11.

**Zhang** H, Yang HC, Yang Y. *PooL: an efficient method for estimating haplotype frequencies from large DNA pool.* Bioinformatics. 2008 Sep 1;24(17):1942-8. Epub 2008 Jun 23.

**Zheng** M, Simon R, Mirlacher M, Maurer R, Gasser T, Forster T, Diener PA, Mihatsch MJ, Sauter G, Schraml P. *TRIO amplification and abundant mRNA expression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer.* Am J Pathol. 2004 Jul;165(1):63-9.